

CHAPTER 7

OBSERVATIONAL COSMOLOGY – MEASURING THE UNIVERSE

7.1 Introduction

Observational cosmology is the branch of science concerned with measuring the parameters that characterize the Universe. Some of the most important of those parameters were introduced in Chapter 5, including the Hubble constant H_0 , the current value of the deceleration parameter q_0 , the age of the Universe t_0 , and the current values of the density parameters for matter and for the cosmological constant (dark energy), $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$, respectively. (The cosmological constant, Λ , can be given as an alternative to $\Omega_{\Lambda,0}$.) The study of the early Universe in Chapter 6, with its emphasis on nuclear and subnuclear processes, introduced several more parameters, including the current value of the density parameter for baryonic matter $\Omega_{b,0}$ (one of the contributions to $\Omega_{m,0}$). Other parameters may arise from any detailed proposal regarding the nature of dark matter. Yet more parameters can be associated with the cosmic microwave background radiation (CMB). Since the CMB has a black-body spectrum, one of the most important of these parameters is the present temperature of the CMB. This is currently one of the best determined of all the cosmological parameters: according to a widely quoted result, its value is $T_{\text{cmb}} = (2.725 \pm 0.002) \text{ K}$.

This chapter is mainly concerned with the determination of just a few of these parameter values, particularly H_0 , q_0 , $\Omega_{m,0}$, $\Omega_{\Lambda,0}$ and $\Omega_{b,0}$. The first four of these determine which of the various Friedmann–Robertson–Walker models (FRW models; introduced in Chapter 5) provides the best description of the Universe we actually inhabit. The fifth, the current value of the density parameter for baryonic matter, represents a particularly familiar contribution to the total matter density. The current ‘best values’ for these five parameters are discussed, along with the methods used to arrive at those values and the degree of uncertainty in each result. As you will see, there is growing confidence among observational cosmologists that they are finally closing in on the true values of these parameters. If this confidence is well placed, it should soon be possible to eliminate much of the uncertainty that has traditionally been associated with determining the most appropriate model for our Universe.

One topic that arises at several different points in our discussion is the CMB. You have already seen that its absolute temperature is known to better than one part in 10^3 , which is an extraordinary level of precision for any cosmological parameter. The precision of this result is an indication of what is now happening throughout observational cosmology. In the view of many cosmologists, we are just entering an age of *precision cosmology*, in which the values of the key cosmological parameters will be determined accurately and precisely. As is explained later, it is our ability to measure and understand the CMB that is leading the way into this new era. However, other methods are still important, since the confidence in precise CMB measurements is based on their agreement with several more traditional kinds of measurement. For this reason, the more traditional methods are discussed in Sections 7.2, 7.3 and 7.4, while Section 7.5 is entirely devoted to the CMB and precision cosmology.

Any account of observational cosmology is likely to become quickly outdated. Even as this chapter was being written, new observational data became available from a space probe called WMAP. These data are discussed in Section 7.5, and give just one indication of the need to be always alert for new findings and new developments in a

subject as vibrant and active as observational cosmology. This chapter can provide nothing more than a snapshot of this fast-moving field.

7.2 Measuring the Hubble constant, H_0

The Hubble constant is in many ways the most fundamental of the cosmological parameters. Of them all, it is the one most easily related to the ‘theoretical’ parameters that characterize the FRW models, and the value of H_0 plays an important part in almost all determinations of the other constants. In this section we discuss two methods of determining H_0 . A third method, based on the CMB, is discussed in Section 7.5. First, though, here are some questions to help you recall what you learned about the Hubble constant and the related Hubble parameter in Chapter 5.

- Briefly describe the curvature parameter k , and the scale factor $R(t)$ that appear in the FRW models, and state which aspects of space–time they describe.
- The *curvature parameter* k may be -1 , 0 or $+1$. It indicates the sign of the curvature of space, and helps to determine geometric properties such as the sum of the interior angles of a triangle, the relationship between the radius and the circumference of a circle, and whether neighbouring straight lines that are initially parallel will converge or diverge. The positive value, $k = +1$, indicates that space has a finite total volume; the values $k = 0$ and $k = -1$ indicate that space is infinite.

The *scale factor* $R(t)$ varies with time, t . It describes the uniform expansion or contraction of space. If the positions of galaxies are expressed in terms of co-moving coordinates (i.e. ones that expand with the Universe) then the coordinate separation of two galaxies might have the fixed value r , but the physical distance between those galaxies (measured in metres, say) will be proportional to $R(t)$, and will increase as $R(t)$ grows and the Universe expands.

- What is the precise relationship between the Hubble constant H_0 and the scale factor $R(t)$? You may find it useful to start by defining the Hubble parameter.
- The *Hubble parameter*, $H(t)$, is a fractional measure of the rate of change of the scale factor (see Section 5.4.1). In terms of symbols, this relationship may be written as

$$H(t) = \frac{\dot{R}(t)}{R(t)} \quad (7.1)$$

where $\dot{R}(t)$ represents the rate of change of R at time t . The *Hubble constant*, H_0 is simply the value of the time-dependent Hubble parameter $H(t)$ at the present time t_0 ; that is to say,

$$H_0 = H(t_0) \quad (7.2)$$

- Write down Hubble’s law and suggest a way in which this can be used to determine the Hubble constant from observations.
- According to Hubble’s law, the redshift (z) of galaxies increases in proportion to their distance (d) from us, provided the redshift is not too great (less than

about 0.2, say). If the constant of proportionality is identified as H_0/c , then Hubble's law may be written

$$z = \frac{H_0}{c} d \quad (7.3)$$

This suggests that if the redshifts of a sample of galaxies are measured, and if the distances of the same galaxies are determined independently of the redshifts, then those observational data can be used to determine H_0/c , and hence H_0 , since c is well known. One way of doing this would be to plot a graph of z against d , determine the best straight line through the data points and then measure the gradient of that line, since this gradient should be H_0/c .

7.2.1 Determinations of H_0 using the Hubble diagram

The first determination of the Hubble constant was by Hubble himself, and was published in 1929. The determination appeared in the paper in which Hubble announced the law that is now named after him, and was based on measurements of the distances of 24 relatively nearby galaxies. The redshifts of these galaxies were already known thanks to the work of Vesto M. Slipher (1875–1969), who had systematically measured the redshifts of almost all the spiral galaxies that were known at the time. Hubble determined the distances of the 24 galaxies by using Cepheid variables and other distance indicators in the manner described in Chapter 2. For those galaxies where Cepheid variables could be seen, the distances were deduced as follows.

- The *period* of each Cepheid was measured by recording the rise and fall of its apparent magnitude and determining the time between successive maxima. Technically this amounts to measuring the *light curve* of the Cepheid – a plot of apparent magnitude against time over a complete period.
- A previously calibrated *period–luminosity relation* for Cepheids was then used to determine the luminosity or (equivalently) the *absolute magnitude* of each Cepheid.
- The absolute magnitude M determined in this way was then combined with a measurement of the apparent magnitude m to determine the *distance* of each of the Cepheids, and this was assumed to be the distance of the galaxy that contained the Cepheid.

Independently of these distance measurements, the formula $v = cz$ was used to assign a ‘naïve’ recession speed v to each of the 24 galaxies on the basis of its measured redshift z . We refer to v as the ‘naïve’ recession speed since the use of $v = cz$ was based on the assumption that the redshift z is the result of a simple Doppler shift, which is not true. The value of v differs from the real recession speed of the galaxy, and the discrepancy increases with the redshift. Nonetheless, the measurements made it possible for Hubble to represent each of the 24 galaxies by a single point on a plot of naïve recession speed against distance, as shown in Figure 7.1.

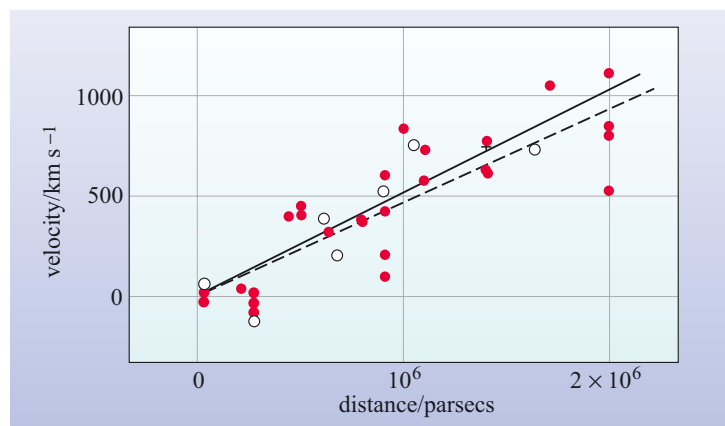


Figure 7.1 The original ‘Hubble diagram’, based on a figure that appeared in 1929. The 24 galaxies are represented by the red dots; the solid line represents the best straight line through the data. The open circles and the dotted line refer to nine groups of galaxies that Hubble also considered. (Courtesy of the National Academy of Sciences)

The data points were somewhat scattered, but Hubble detected a linear trend, and on this basis claimed that the recession speed of galaxies increased in proportion to their distance from us. From the gradient of the straight line in Figure 7.1, Hubble deduced that the rate of increase of recession speed with distance was $500 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Using the notation and terminology of modern astronomers, the relation that Hubble proposed can be rewritten as $v = H_0 d$, where $v = cz$, and H_0 is Hubble's constant. From this we obtain the modern form of Hubble's law, $z = (H_0/c)d$, and we can see that $500 \text{ km s}^{-1} \text{ Mpc}^{-1}$ represents the first determination of the Hubble constant H_0 . This value is far outside the range of values that are currently thought to be credible, but the measurement is historically important as the start of a long campaign to measure H_0 .

There have now been hundreds of attempts to determine the Hubble constant. Like all observational results, these determinations are subject to various kinds of uncertainty. There are, for instance, the **random uncertainties** that beset any measurement, and which generally cause different determinations of a measured quantity to vary about some mean value. In addition there may be **systematic uncertainties**, which will always influence the measured value in the same way, no matter how many times the measurement is repeated. It is now known that there were substantial systematic uncertainties in the early determinations of the Hubble constant. Some of these were the result of studying only nearby galaxies where the motion was dominated by 'local' effects, such as the attraction of the local supercluster (see Chapter 4), rather than the large-scale Hubble flow. Others arose from the mistaken belief that two different kinds of variable star were actually the same kind of Cepheid. However, even after these sources of error were removed, different determinations of the Hubble constant still tended to disagree by substantial amounts. This is indicated in Figure 7.2, where the results of various determinations of H_0 are plotted against the year in which they were made. Note that uncertainty ranges, represented by vertical bars through the data points, have been used to show the range of values that are consistent with the measured results, rather than pretending that each measurement results in a unique value. Estimating the appropriate size for the uncertainty range in an astronomical

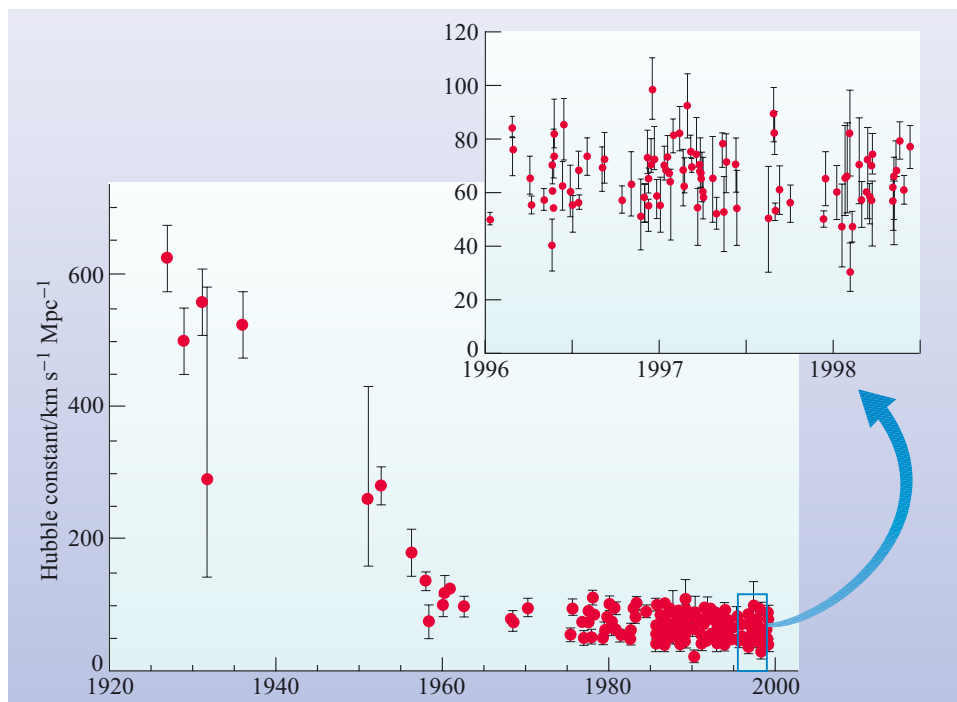


Figure 7.2 The evolution of the measured value of the Hubble constant H_0 . Over 300 measurements performed since 1975 have yielded values between 50 and $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. (Adapted from *Sky and Telescope*, based on work by J. Huchra and the HST Key Project on the Distance Scale)

measurement is often highly complicated and very time consuming, but it is vital if other scientists are to properly appreciate the significance of the results.

As the figure shows, by the 1980s a large number of determinations indicated that the Hubble constant was probably somewhere in the range between 50 and 100 km s⁻¹ Mpc⁻¹, and for a while there were strong proponents of either end of that range. Perhaps the most notable were Gerard de Vaucouleurs who favoured a high value of the Hubble constant, and Allan Sandage who favoured a low value. The disagreement between these eminent astronomers was not caused by the inherent difficulty of making the observations, but rather by the different procedures they adopted when interpreting their observations and correcting them for various observational effects.

In an effort to reduce the uncertainty, one of the designated ‘*key projects*’ of the Hubble Space Telescope (HST) in the 1990s was a programme to determine the Hubble constant to an accuracy of about 10%. The efforts of a team of astronomers, led by Canadian-born US astronomer Wendy Freedman, resulted in the value $H_0 = 72 \pm 8$ km s⁻¹ Mpc⁻¹, which has now been supported by more recent CMB-based measurements.

The basis of the HST Key Project to measure H_0 was to use the Cepheid period–luminosity relationship to calibrate other methods of distance measurement. The period–luminosity relation that was used was based on studies of Cepheids in our neighbouring galaxy, the Large Magellanic Cloud (LMC), which is sufficiently close for large numbers of Cepheids to be observed. The calibration of the Cepheid period–luminosity method required the distance to the LMC to be known, and it was assumed, on the basis of measurements made by other researchers, that the LMC is at a distance of 50 kpc from the Sun.

The first stage of observations was to search for Cepheids in selected galaxies which lie between about 3 and 25 Mpc from the Milky Way. At these distances the ‘local’ motion of galaxies still has a significant effect on their measured redshifts, so the selected galaxies could not themselves be used to determine the Hubble constant. However, once the Cepheids had been found and measured, the distances of the selected galaxies could be precisely determined. Once this had been done, a variety of measurements made in those selected galaxies could be used to calibrate other methods of distance measurement that could be applied to more remote galaxies, which were more likely to represent the Hubble flow. The galaxies that were searched for Cepheids were carefully chosen so as to allow the calibration of not one, but five other methods of distance measurement, all of which were discussed in Chapter 2.

■ On the basis of Chapter 2, name five methods of galactic distance determination (excluding the use of Hubble’s law) that are appropriate for measuring distances to about 100 Mpc. Briefly explain the nature of the ‘calibration problem’ that these methods must confront.

□ The five methods appropriate for measuring distances to about 100 Mpc are: the Type Ia supernova method; the Type II supernova method; the methods based on the Tully–Fisher and fundamental plane relations; and the method based on surface brightness fluctuations.

The calibration problem consists of making measurements in relatively nearby galaxies at known distances (between about 3 and 25 Mpc in the case of the HST Key Project) that enable the ‘relative’ distances indicated by standard candles in more remote galaxies to be converted into ‘absolute’ distances that may be expressed in units such as megaparsecs.

Table 7.1 A summary of the different methods used to measure H_0 from the HST Key Project.

Distance measurement method	Number of Cepheid calibration measurements	Number of measurements used to obtain H_0	Range of redshift in the measurements used to obtain H_0	Value of H_0 obtained/ $\text{km s}^{-1} \text{Mpc}^{-1}$
Type Ia supernovae	6	36	0.013–0.10	71
Type II supernovae	4	4	0.006–0.047	72
Tully–Fisher relation	21	21	0.003–0.030	71
fundamental plane	3	11	0.003–0.037	82
surface brightness fluctuations	6	6	0.013–0.019	70

The five methods used to determine H_0 , along with their results, are shown in Table 7.1. The numbers of measurements used to calibrate the five methods are shown in the second column. The fact that some techniques were calibrated using a small number of galaxies highlights the difficulty of the calibration problem: prior to the HST measurements, some of the methods that are listed had never before been calibrated against the Cepheid period–luminosity method.

Having calibrated the five methods, the next stage was to use those methods to determine the distances of galaxies that are sufficiently far away that their measured redshifts are dominated by the expansion of the Universe rather than by any ‘local’ effects. The typical uncertainty in redshift caused by local effects is roughly 10^{-3} , so in order to bring this uncertainty to below 10% of the measured redshift, that measured redshift must be at least $z \approx 10^{-2}$. The actual ranges in redshift that were sampled using the five different methods of distance determination are shown in Table 7.1, along with the number of measurements that were used to obtain a value of H_0 .

The correlation between redshift and distance for a number of galaxies and clusters, based on the HST Key Project results, is shown in Figure 7.3. One of the remarkable outcomes, evident from the fifth column of Table 7.1, is that the five different methods, which are based on quite different physical principles, agree well with one another. The values obtained for H_0 vary from 70 to $82 \text{ km s}^{-1} \text{Mpc}^{-1}$. This level of agreement provides some reassurance that there are no gross systematic differences

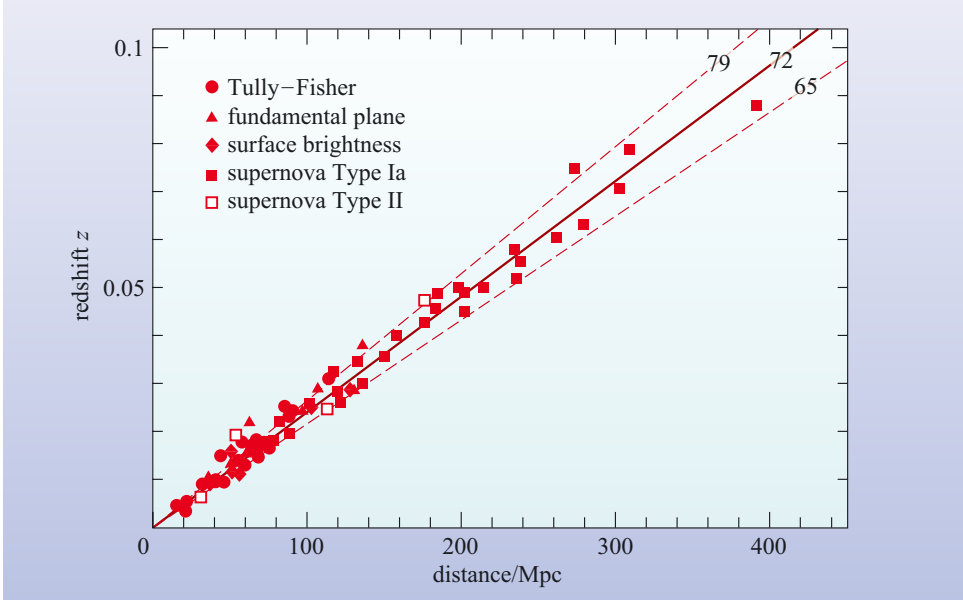


Figure 7.3 A modern version of the Hubble diagram from the HST Key Project, with recession speed replaced by redshift. Note that in this case the gradient of the graph will be equal to the quantity H_0/c . Various straight lines through the data are shown, together with the corresponding values of H_0 measured in $\text{km s}^{-1} \text{Mpc}^{-1}$. (Adapted from Freedman *et al.*, 2001)

between the five methods that were adopted, and leads, after further analysis, to the result quoted earlier: $H_0 = 72 \pm 8 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

Plots such as those in Figures 7.1 and 7.3 are examples of what are generally called **Hubble diagrams**. They still play an important part in the determination of the Hubble constant, although in modern versions, such as Figure 7.3, it is often the redshift z that is plotted along the vertical axis (rather than the ‘naive’ recession speed) and the range of redshifts is generally much greater than in the original Hubble diagram of Figure 7.1. This last point is important since it is only by extending the Hubble diagram to sufficiently distant galaxies, as the HST Key Project did, that there is any hope of escaping the effects of local motion and determining the true rate of expansion of the Universe.

7.2.2 Other methods of determining H_0

Because it has been so difficult to obtain convergence amongst measurements of H_0 based on the use of the Hubble diagram, there has always been an interest in other methods that might be used to determine the Hubble constant. A method based on detailed observations of the cosmic microwave background radiation is discussed in Section 7.5. Yet another method involves the phenomenon of *gravitational lensing* that was introduced in Chapter 4.

You will recall from Chapters 4 and 5 that, according to general relativity, matter and radiation cause space–time to curve, and the path of a light ray responds to that space–time curvature. This means that concentrations of matter, such as galaxies or clusters of galaxies, can act as ‘gravitational lenses’, which can magnify and distort an observer’s view of objects that lie beyond the lens. A possible outcome of gravitational lensing is the formation of multiple images of the kind shown in Figure 7.4. In this case, the gravitational lens consists of a large, elliptical galaxy and a surrounding cluster of galaxies, while the two-part image provides two views, A and B, of a single quasar, QSO 0957+561 (quasars and their variable luminosities were discussed in Chapter 3). The redshift of this quasar is 1.41, while that of the elliptical galaxy is 0.36. The way the lens forms this image is shown in Figure 7.5.

As Figure 7.5 indicates, the two images of the quasar reach the observer by different pathways that have different lengths. This means that any change that occurs in the quasar, such as a sudden brightening, will be observed in one image some time after it is observed in the other. In other words, as the two images fluctuate

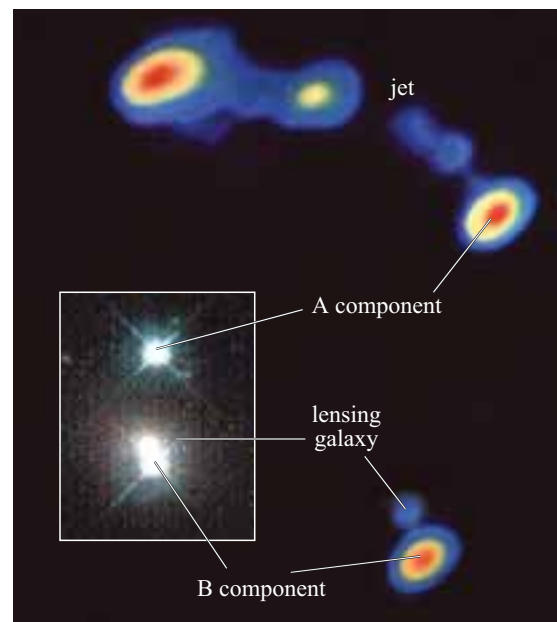
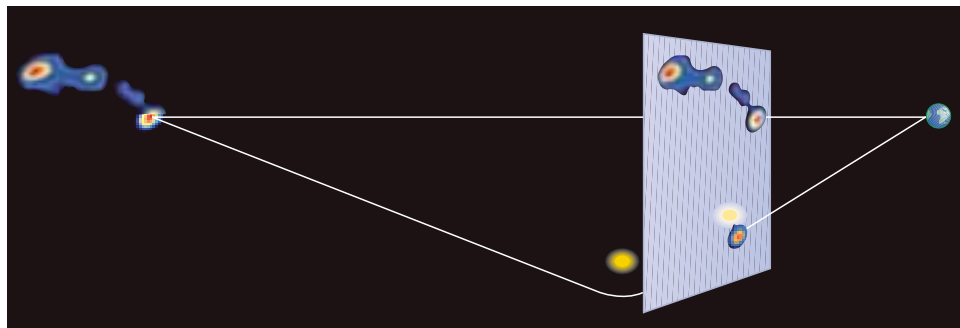


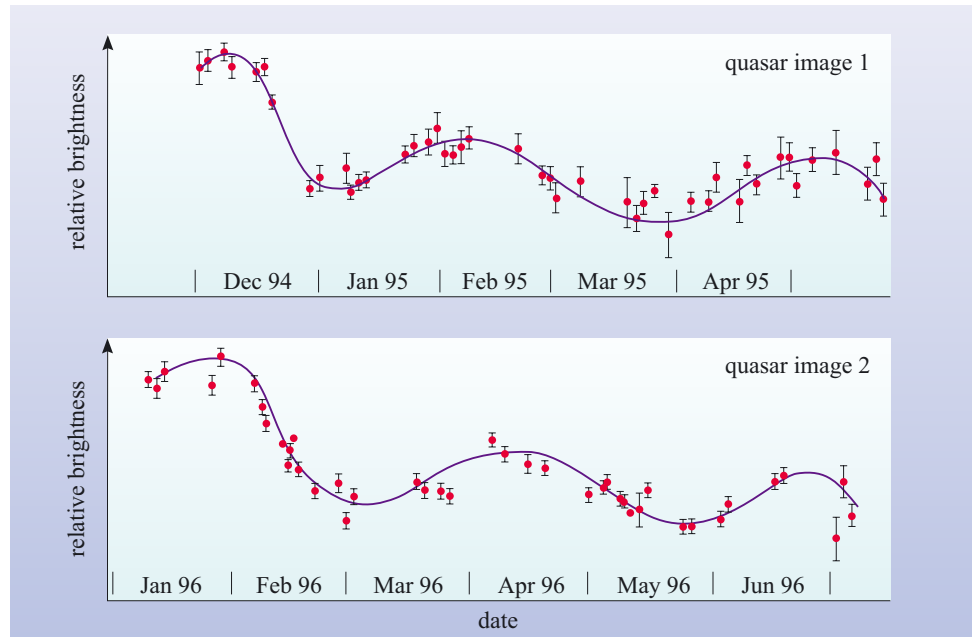
Figure 7.4 The two-component image of the gravitationally lensed quasar QSO 0957+561. The large, elliptical galaxy that is partly responsible for the lensing can be seen just above the B component of the quasar image. The main image was recorded by a radio telescope, the inset is an optical image. The jet is part of the quasar. (NRAO/AUI)

Figure 7.5 A schematic diagram, with an exaggerated angular scale, showing how the double image of QSO 0957+561 is formed. The distance to the lensing galaxy is a little over 1000 Mpc, about one-third of the distance to the quasar. Note that the two routes from the quasar to the observer are not of equal length. (Adapted from Henbest and Marten, 1983)



in response to changes in the quasar itself, there will be some time lag between the observed patterns of fluctuation of the two images due to the different path lengths from the quasar. Unfortunately, the difference in path length is not the only cause of a time lag between fluctuations in the two images. The distribution of material within the gravitational lens determines the detailed geometry of space in the neighbourhood of the lens, and this influences the time of passage for light as it passes through the gravitational lens. Light passing through the denser part of the lens is generally delayed relative to light passing through the less dense parts. Since the light responsible for the two images passes through different parts of the lens, just such a density-dependent time-of-passage contribution to the observed time lag between fluctuations in the two images can be expected. The total time lag, including both the path-length and the time-of-passage effect, has been measured for QSO 0957+561 and for a few other gravitational lenses (see Figure 7.6). In the case of QSO 0957+561 the time lag is about 415 days.

Figure 7.6 The observed patterns of brightness fluctuation in the two images of a double quasar differ by a fixed time lag. (Adapted from Schild and Thomson, 1997)



In the 1960s, well before the discovery of the first gravitational lens, it was known from theoretical calculations that the path length contribution to the time delay would depend on the Hubble constant. So, in principle, this part of the observed total time lag can be used to determine the Hubble constant without the need for calibration. However, in order to carry out such a determination it is first necessary to calculate the time-of-passage delay and to subtract that from the observed total time lag. Performing such a calculation involves formulating a theoretical ‘model’ of the lens and making a number of assumptions about the way that gravitating material is distributed within the lens. The need to model the lens in this way introduces a large degree of uncertainty into what might otherwise be a rather precise way of evaluating the Hubble constant. In the particular case of QSO 0957+561, the lens is dominated by a large, elliptical galaxy that is fairly simple to model. Even so, there are still several remaining uncertainties. At the time of writing, the best current estimate of H_0 from this particular source is about $(61 \pm 6) \text{ km s}^{-1} \text{ Mpc}^{-1}$, but this is about 30% different from the ‘best’ value provided by the same source just a few years ago – the difference being almost entirely due to changes in the modelling.

Because of this sensitivity to modelling assumptions it's probably best to regard the gravitational lensing technique as providing a useful cross-check on the broad value of the Hubble constant, but to still need considerable improvement (at least in the certainty of theoretical modelling) before it can match other methods.

QUESTION 7.1

Listed below are some possible sources of error and uncertainty in a determination of the Hubble constant based on gravitational lensing. Classify each as either random or systematic, justifying your decision in each case.

- (a) Uncertainty about the distribution of dark matter in the lensing galaxy.
- (b) Failure to detect a small companion galaxy located directly behind the lensing galaxy.
- (c) Difficulty, due to incomplete data, in determining the total time lag between fluctuations in the two images of the lensed galaxy.

7.3 Measuring the current value of the deceleration parameter, q_0

In the context of the Friedmann–Robertson–Walker models, the Hubble constant, H_0 , measures the rate of expansion of the Universe. However, the Hubble constant only represents the *current* rate of expansion: the *current* value of the more general Hubble parameter $H(t)$ that describes the rate of expansion at *any* time t . The rate of expansion might have been higher or lower in the past, and may, in principle at least, be greater or smaller in the future. The observable quantity that tells us whether the cosmic expansion rate is currently increasing or decreasing is the current value of the deceleration parameter q_0 , the current value of the time-dependent deceleration parameter $q(t)$ that was introduced in Chapter 5. A positive value of q_0 would indicate that the expansion is slowing down, a negative value that it is speeding up.

The current value of the deceleration parameter is expected to influence the relationship between the distance and the redshift of distant galaxies. As you saw in Chapter 5, to a first approximation this relationship is described by Hubble's law:

$$d = \frac{cz}{H_0}$$

but a more accurate relationship, according to the Friedmann–Robertson–Walker models, is given by

$$d = \frac{cz}{H_0} \left[1 + \frac{1}{2}(1 - q_0)z \right] \quad (7.4)$$

As was shown in Figure 5.29, these two different expressions agree well at low redshift, but they differ significantly when z is greater than about 0.2.

Equation 7.4 provides a way of determining the current value of the deceleration parameter from observation. It implies that a plot of d against z for distant galaxies will show systematic deviations from the straight line implied by Hubble's law if q_0 has any value other than $q_0 = 1$. In principle, then, we can expect to determine q_0 by

making accurate independent measurements of d and z and then determining which values of H_0 and q_0 in Equation 7.4 provide the best fit to the observed data. In practice, of course, this procedure is nothing like as simple as it sounds. You saw in the last section the enormous difficulty that astronomers have had in using this kind of approach to pin down the value of H_0 , without the added complication of determining q_0 as well. With this in mind, you will not be surprised that many of the early attempts to measure q_0 were not particularly successful. Nonetheless, they deserve a brief discussion because of the light they shed on the observational challenges of measuring cosmic deceleration.

7.3.1 Early attempts to determine q_0

The direct proportionality between redshift and distance indicated by Hubble's law is in excellent agreement with the data obtained from galaxies out to redshifts of 0.1 or so. Hence any attempt to detect departures from Hubble's law must involve galaxies with redshifts greater than 0.1. Such measurements were not really possible until the late 1940s, when the 200-inch telescope on Mount Palomar (see Figure 7.7) was commissioned. Even then, observing individual bright stars in such distant galaxies was not possible, so distance determinations had to make use of clusters of stars, or even whole galaxies.

Early attempts to measure deceleration using the Palomar telescope concentrated on whole galaxies. Hubble had pointed out in 1936 that, if galaxies had some natural upper limit to their brightness, then the brightest galaxies in clusters that contained hundreds or thousands of galaxies might be expected to be close to that limit. Thus the brightest galaxy in a cluster might represent a 'standard candle' (i.e. a source of fixed luminosity or absolute magnitude), and once that absolute magnitude had been determined it should be possible to work out the distance of any such galaxy from an observation of its apparent magnitude. (This was the basis of the technique used by George Abell to determine the distances of the rich clusters discussed in Chapter 4.) Accepting this idea, an extensive survey involving clusters with redshifts up to $z = 0.18$, published in 1956, found evidence of curvature in the Hubble diagram and led to the value $q_0 = 3.7 \pm 0.8$. However, another study soon gave $q_0 = 1.0 \pm 0.5$, and a reconsideration of earlier work led to a claim that $q_0 = 0.2 \pm 0.5$. Clearly, this approach was not yielding consistent results, although there did seem general agreement that q_0 was probably positive, as might be expected in a Universe where gravity gradually slowed the cosmic expansion. By the late 1970s extensive studies of galaxies, theoretical as well as observational, were showing that attempts to use the brightest galaxies in clusters as standard candles were being seriously undermined by evolutionary effects, since the galaxies themselves showed signs of changing over time. The larger the redshift of the galaxy, the greater the time its light had spent travelling to the Earth, and the 'younger' the galaxy in cosmic terms; these youthful, bright galaxies did not, it was becoming clear, have the same luminosity as their more evolved 'older' counterparts at lower redshifts. Since these evolutionary effects were overwhelming the effects due to cosmic deceleration, the brightest-galaxy approach became discredited and some other technique had to be used to determine q_0 .



Figure 7.7 The 200-inch telescope at the Mount Palomar observatory. The large, white structure that supports the open lattice-work of the telescope itself is aligned with the Earth's rotation axis to facilitate the tracking of stars. (Caltech Archives)

The alternative that came to the fore was based on the use of *supernovae*. As you know from earlier discussions, supernovae are highly luminous events that mark the explosive death of certain kinds of star. They can be classified into various types according to their spectra, with Type I supernovae being the most luminous. Astronomers are now keenly aware of the existence of various subclasses of Type I supernovae, such as the Type Ia supernovae that were discussed in Chapter 2. However, the early attempts to use supernovae to determine q_0 treated all Type I supernovae on the same basis.

Type I supernovae can easily be observed in galaxies with redshifts that are greater than 0.1, and may even be detected at redshifts as great as $z = 1$. This ensures that the galaxies containing the more distant supernovae are so far away that they are unlikely to be significantly influenced by local disruptions of the Hubble flow. Hence any value of q_0 based on observations of distant supernovae might be expected to be free of the systematic uncertainties that might affect more ‘local’ measurements.

Making the assumption that all Type I supernovae attain the same maximum luminosity, attempts were made to use them to determine q_0 . The basic approach was similar to that based on brightest galaxies in clusters: measurements of the supernova’s apparent magnitude were combined with the assumed ‘standard’ value of the absolute magnitude to determine the distance, while a separate measurement yielded the supernova’s redshift. The way that the redshift varied with distance, for an appropriate sample of Type I supernovae, was expected to reveal the value of q_0 . Unfortunately this particular approach was recognized as unreliable when it was realized that the broad class of Type I supernovae actually included several quite distinct types of supernova that differed in intrinsic brightness. However, this failure paved the way for the more recent attempts to measure the current value of the deceleration parameter using the carefully studied subclass of supernovae known as Type Ia supernovae. The results of this study have had such a significant impact on observational cosmology that they deserve to be discussed in a separate section.

QUESTION 7.2

Use Hubble’s law to calculate the distance of a supernova in a galaxy with redshift $z = 1.0$. Explain why the distance you have estimated may not be a good estimate of the true distance to the galaxy. How does the distance you calculated compare with the size of the local supercluster, as discussed in Chapter 4?

7.3.2 Determinations of q_0 using Type Ia supernovae

Type Ia supernovae are believed to occur when a white dwarf star in a close binary system accretes so much matter from its binary companion that it exceeds a critical mass (about 1.4 times the mass of the Sun) and collapses under its own weight. In a rapid succession of events the collapse is transformed into an explosion that causes a rapid brightening of the supernova, followed by a more gradual decline that can typically be observed for a month or more.

Because the mass required to initiate the supernova is thought to be the same in all cases, and because the physical processes involved are believed to be always very similar, it is expected that all Type Ia supernovae attain approximately the same maximum brightness. This means that Type Ia supernovae might be used as ‘standard candles’ just like Cepheid variables with a known period.

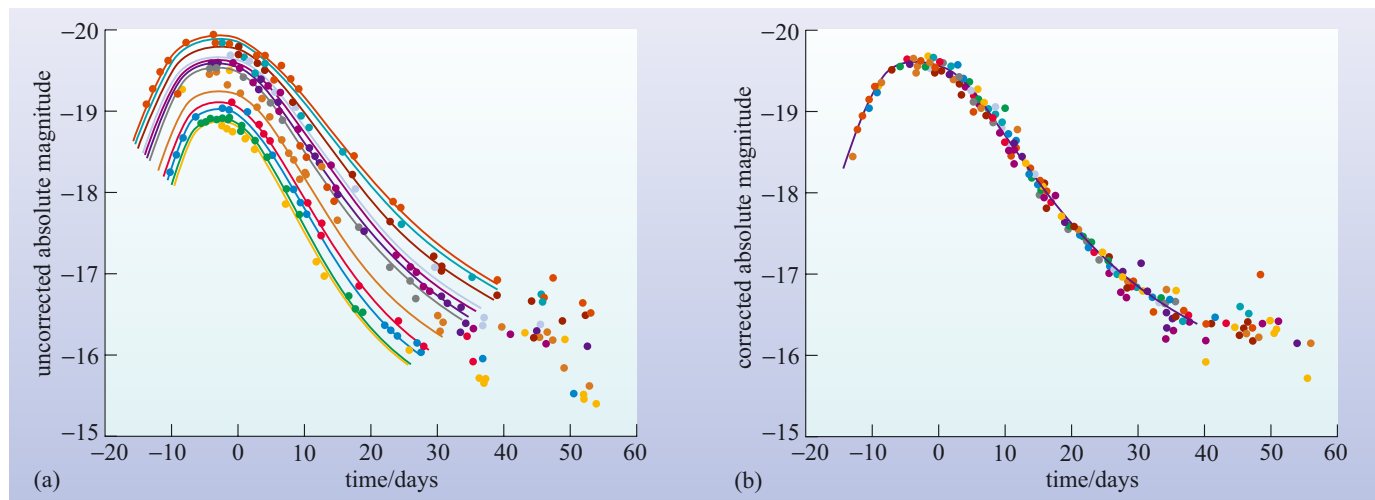


Figure 7.8 The light curves in (a) show the absolute magnitude of relatively nearby Type Ia supernovae as a function of time. Note that the greater the maximum luminosity the slower the decline in brightness from that maximum. (Crudely, the higher the curve, the wider it is.) Part (b) shows the effect of ‘correcting’ all the light curves according to a standard prescription derived from the observations. The ‘corrected’ light curves show much less intrinsic spread in maximum luminosity. (Adam Reiss)

In practice, simply expecting all Type Ia supernovae to have approximately the same total luminosity is not a sufficiently good basis for a major observational research programme. Such a programme must also be supported by an extensive and painstaking investigation of Type Ia supernovae that enables any intrinsic differences between Type Ia supernovae to be studied, understood and, as far as possible, eliminated as a source of uncertainty. This constitutes the process of *calibration* that was described in general terms in Chapter 2. In this case it can be achieved by studying Type Ia supernovae that occur in galaxies whose distances can be determined independently of the supernovae. This work was effectively carried out in the early 1990s by a number of investigators, each pursuing their own goals. As a result, it was found that even within the tightly defined subclass of Type Ia supernovae there were still differences of around 35% in the maximum luminosity attained by different Type Ia supernovae. However, it was also found that the value of the maximum brightness correlates with the rate at which the brightness of the supernova declines from its maximum (see Figure 7.8). Thanks to this correlation it became possible to ‘correct’ the observed brightness of any particular Type Ia supernova in such a way that it behaved as though its maximum luminosity was actually within 15 to 20% of the maximum luminosity of any other Type Ia supernova. Unfortunately this correction lacks a sound theoretical basis, but it does appear to be what is needed to substantially reduce the intrinsic differences in luminosity between different Type Ia supernovae. Calculations indicated that ‘corrected’ observations of thirty or so Type Ia supernovae with redshifts roughly in the range 0.5 to 1 should provide a good indication of cosmic deceleration.

- In general terms, what sort of observations would be necessary in order to ‘correct’ the brightness of any particular Type Ia supernova with the aim of minimizing the effect of intrinsic differences in luminosity?
- The apparent magnitude of the supernova would have to be measured repeatedly over a sufficiently long period (about a month) to determine the rate of decline of its light curve, and hence the supernova’s peak luminosity relative to other Type Ia supernovae.

By the mid-1990s Type Ia supernovae were sufficiently well understood that they could reasonably be used to determine the current value of the deceleration parameter, and two independent teams of researchers set out to do this. The two

groups – the High- z Supernova Search Team and the Supernova Cosmology Project – faced similar challenges. Each group had to detect a reasonable number of Type Ia supernovae, measure the redshift of each, and carry out sufficient follow-up observations over a period of several weeks to determine the rate of decline in brightness of each supernova. They would then be able to ‘correct’ their observations in order to minimize the effect of the intrinsic differences in maximum brightness that exist between different Type Ia supernovae.

This programme presents many observational challenges and requires the involvement of several astronomers using a range of telescopes and detectors. The High- z Supernova Search Team, for example, included about two dozen astronomers from fifteen different institutions spread over four continents. Their search for distant Type Ia supernovae involved recording large-area images of the sky with sufficient sensitivity to detect the faint images of supernovae out to redshifts of about $z = 1$. This was initially done using large-area CCD detectors on the 4-metre telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile, where the team was able to image about three square degrees of sky per night, down to a limiting magnitude of about 23. In order to identify candidate supernovae they imaged the same area of sky 21 days later and then compared the two images, looking for point-like objects that had changed in brightness over that time. (The period of 21 days was chosen to reflect the ‘rise time’ of a typical Type Ia supernova – the time it takes for the light curve to build up to a maximum, before starting its much slower decline.) The comparison was mainly an automatic process in which a computer program aligned the two images, compensated for various observational effects, and eliminated known stellar sources in our own Galaxy before presenting the astronomers with the data from which to make a final selection of candidate Type Ia supernovae. This resulted in about five to 20 candidates for each night of observation, most of which were later confirmed as Type Ia supernovae.

Once the candidates had been identified the detailed follow-up work could begin. This involved studying each candidate individually, both *spectroscopically* (to determine its redshift and to confirm that it really was a Type Ia supernova), and *photometrically* (to determine the light curve and hence the ‘corrected’ maximum brightness of the source). In the case of the High- z Supernova Search Team, the spectra were obtained using one of the 10-metre Keck telescopes in Hawaii (Figure 7.9). A very large telescope was needed because many of the candidates were very faint, so recording the light in their spectra required a large-aperture telescope that could capture as much light as possible in a relatively short time. The photometric (brightness) studies were carried out using a range of telescopes in both the northern and southern hemispheres. For some of the sources, even the Hubble Space Telescope was used. Accurate photometry generally involves comparing the object being studied with stars whose brightness has already been determined with great care, and often requires the observations to be made through a set of standard filters designed to pass light in specific ranges of wavelength. The use of filters allows the effects of obscuration (by dust in the host galaxy) to be determined and eliminated.

It’s important to note that the procedure adopted by the High- z Team was not designed to accurately determine the absolute magnitude of the high redshift supernovae. Rather, the aim was to

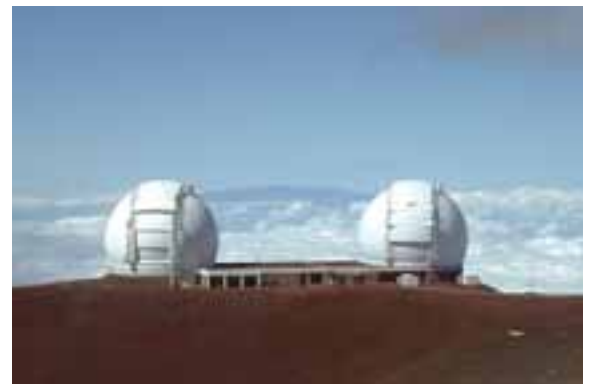


Figure 7.9 The twin domes of the two 10-metre Keck telescopes in Hawaii. (WY’east Consulting, 2002)

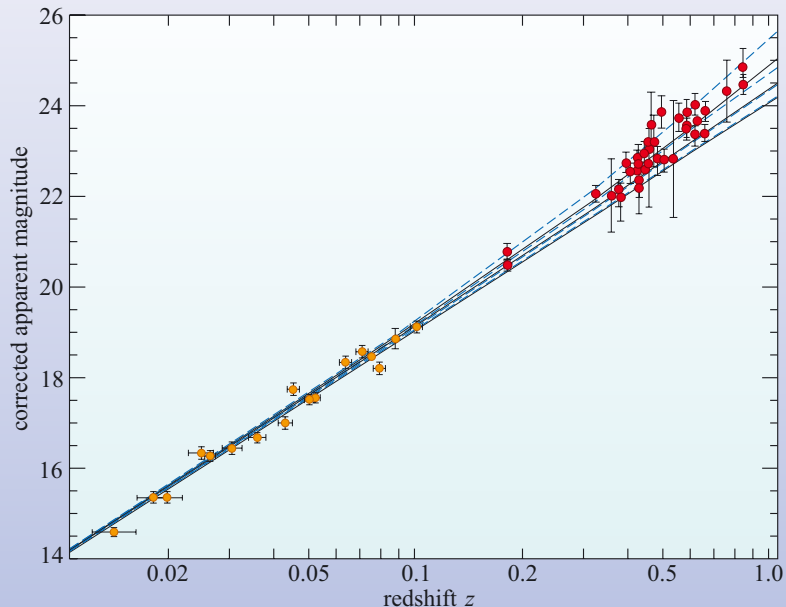
determine their brightness relative to one another (and relative to lower redshift Type Ia supernovae). This approach has the advantage of being simpler than methods requiring accurate measurements of absolute magnitudes, while still allowing deviations from Hubble's law to be detected.

The results from both the High- z Team and from the Supernova Cosmology Project began to appear in 1998. In both cases the crucial data took the form of a graph of 'corrected' apparent magnitude against redshift (see Figure 7.10), and indicated that distant Type Ia supernovae have smaller redshifts than would be expected on the basis of observations of nearer Type Ia supernovae and Hubble's law. Amazingly, the results actually imply that the cosmic expansion is accelerating. So, in contrast to all the early determinations of q_0 discussed in the last section, it now appears that the current value of the deceleration parameter is negative, probably in the range $-0.6 < q_0 < 0$.

More recent results from various sources have tended to confirm these initial findings, but, given the history of measurements in this field, there is still considerable concern about the reliability of the results. The quality of the work that has been done is not doubted, but the possibility exists that systematic uncertainties may have influenced the results.

- On the basis of what you learned in the last section, suggest one possible source of systematic error that might influence the Type Ia supernova measurements.
- Evolution. Just like the brightest galaxies in clusters, the more distant Type Ia supernovae might differ in some significant way from their nearer counterparts at smaller redshift. This factor has been considered, but it is hard to know what its precise effect might be.

Figure 7.10 A plot of apparent magnitude against redshift for Type Ia supernovae that have been 'corrected' for various factors, including the intrinsic differences in brightness indicated by the differing rates of decline of their light curves. The 40 red dots represent observations by the Supernova Cosmology Project. The yellow data points represent the nearer supernovae that were used to calibrate the Type Ia light curve. The various lines shown on the graph correspond to different values of q_0 . (Adapted from Schwarzschild, 1998, based on the work of S. Perlmutter *et al.*)



Further confirmation of cosmic acceleration has come from the observation of individual, very distant, supernovae such as one found in the Hubble Deep Field at a redshift of 1.7. These too indicate that the rate of cosmic expansion is currently greater than it has been in the past.

Despite the remaining concerns, the Type Ia supernova results have led to a major reassessment of the viability of the various FRW models. An accelerating cosmic expansion implies a non-zero cosmological constant, and in such a situation the cosmological implications of the Type Ia supernova measurements are not well described by simply quoting a value for q_0 . The results are better described in terms of their implications for various contributions to the cosmic density, so it is to measurements of those quantities that we now turn.

QUESTION 7.3

Studies of Type Ia supernovae have been used in the attempt to measure the Hubble constant, but the high redshift measurements detailed in this section are not appropriate for this purpose. Why not? (*Hint*: look at the quantities that have been plotted in Figure 7.10.)

7.4 Measuring the current values of the density parameters $\Omega_{\Lambda,0}$, $\Omega_{m,0}$ and $\Omega_{b,0}$

Chapters 5 and 6 introduced a number of density parameters that play an important role in characterizing the contents of the Universe. Each of these parameters was defined as the ratio of some kind of density to the critical density $\rho_{\text{crit}}(t)$, where $\rho_{\text{crit}}(t) = 3H^2(t)/8\pi G$ represents the density in a ‘critical Universe’ with Hubble parameter $H(t)$ where both the cosmological constant and the curvature parameter are zero (i.e. $\Lambda = 0$ and $k = 0$). The three density parameters that will be mainly of concern in this section are:

the density parameter for matter

$$\Omega_m(t) = \frac{\rho(t)}{\rho_{\text{crit}}(t)} \quad (7.5)$$

the density parameter for baryonic matter

$$\Omega_b(t) = \frac{\rho_b(t)}{\rho_{\text{crit}}(t)} \quad (7.6)$$

the density parameter for the cosmological constant

$$\Omega_\Lambda(t) = \frac{\rho_\Lambda}{\rho_{\text{crit}}(t)} \quad (7.7)$$

where $\rho(t)$ represents the average cosmic density of all kinds of matter (baryonic and non-baryonic), $\rho_b(t)$ represents the density of baryonic matter, and ρ_Λ has no simple physical interpretation but is probably best regarded as a convenient way of representing the cosmological constant, since $\rho_\Lambda = \Lambda c^2/8\pi G$. (You will recall that Ω_Λ is also referred to as the density parameter for dark energy.) Note that all three of these density parameters, $\Omega_\Lambda(t)$, $\Omega_b(t)$ and $\Omega_m(t)$ depend on time, since $\rho_{\text{crit}}(t)$, $\rho_b(t)$

and $\rho(t)$ all vary with time (though ρ_Λ does not). As in earlier sections, our concern is with determinations of the *current* values of these time-dependent parameters, which we denote $\Omega_{\Lambda,0}$, $\Omega_{b,0}$ and $\Omega_{m,0}$.

7.4.1 Constraints on $\Omega_{\Lambda,0}$ and $\Omega_{m,0}$ from Type Ia supernovae

In the context of the FRW models, the deceleration parameter is related to a combination of density parameters (see Section 5.4.3). If the very small contribution due to radiation is ignored, the Friedmann equation implies

$$q(t) = \frac{\Omega_m(t)}{2} - \Omega_\Lambda(t) \quad (7.8)$$

As this suggests, the results of both the Supernova Cosmology Project and the High- z Supernova Team can be expressed in terms of constraints on the current values of $\Omega_\Lambda(t)$ and $\Omega_m(t)$. In fact, this is the best way of presenting those results, and both research groups chose to present their findings in this way. The results obtained by the two groups are very similar: those of the Supernova Cosmology Project are shown in Figure 7.11. On a plot of $\Omega_{\Lambda,0}$ against $\Omega_{m,0}$ the results appear as a set of ellipses, marked with figures that represent various **confidence levels**. This is a way of indicating the uncertainties that attend any experimental or observational result. The mauve ellipse showing the 99% confidence level, for instance, indicates that, given the measured results, there is a 99% chance that the true values of $\Omega_{\Lambda,0}$ and $\Omega_{m,0}$ are located within the contour. The narrower contours represent increasingly tight constraints on the values of $\Omega_{\Lambda,0}$ and $\Omega_{m,0}$ and are presented with decreasing levels of confidence. All these confidence levels implicitly assume that all sources of uncertainty have been properly taken into account. In the case of the supernova results there is, of course, great concern about the possibility of unrecognized systematic uncertainties, possibly due to evolution, but also possibly due to some other cause. The blue dashed curves indicate the observers' estimates of the 'worst case' implications of various systematic uncertainties, but concerns still remain.

As you can see, the results of supernova cosmology are consistent with a whole range of values for $\Omega_{\Lambda,0}$ and $\Omega_{m,0}$. A significant feature of these results however, is that they strongly exclude a FRW model in which $\Omega_{\Lambda,0} = 0$ and $\Omega_{m,0} = 1$, which was exactly the kind of model Universe favoured by many cosmologists until the supernova data were published.

Although the supernovae results alone do not determine $\Omega_{\Lambda,0}$ and $\Omega_{m,0}$ very precisely, the results have been used in conjunction with the outcomes of other studies to deduce tight constraints on these two values. Some of these additional studies are described later, but they broadly support the idea that the Universe has a flat spatial geometry, implying that $k = 0$ and, consequently, $\Omega_{\Lambda,0} + \Omega_{m,0} = 1$. In the context of Figure 7.11 this implies that the point representing the true values of $\Omega_{\Lambda,0}$ and $\Omega_{m,0}$ lies on the diagonal line marked 'flat geometry'. Accepting this and working to a reasonable level of confidence, the results of the Supernova Cosmology Project indicate that

$$\Omega_{m,0} = 0.28 \pm 0.09 \quad (7.9)$$

while the High- z Supernova Team find

$$\Omega_{m,0} = 0.32 \pm 0.10 \quad (7.10)$$

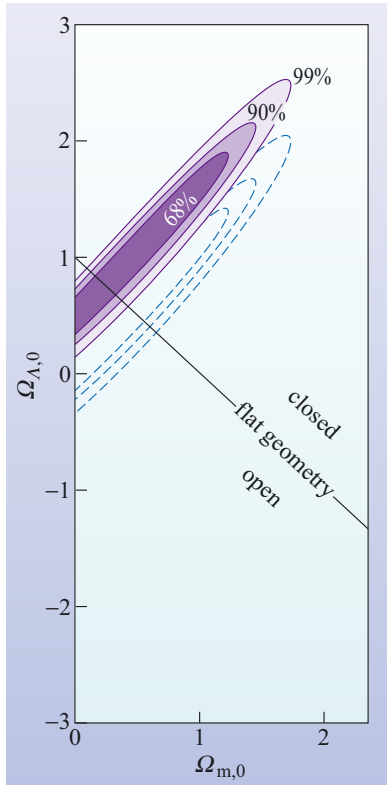


Figure 7.11 Results of the Supernova Cosmology Project, plotted as constraints (indicated by confidence level) on the current values of Ω_Λ and Ω_m . The results effectively rule out the kind of Universe in which $\Omega_{\Lambda,0} = 0$ and $\Omega_{m,0} = 1$ that was favoured by many cosmologists prior to the publication of the supernova data. (Adapted from Schwarzschild, 1998, based on the work of S. Perlmutter *et al.*)

The agreement between these results (within the quoted uncertainties) is impressive, and while there must always be concerns about systematic uncertainties, there is now widespread agreement that $\Omega_{m,0}$ is within the range indicated by these results. Further evidence of this is presented in Section 7.5.

QUESTION 7.4

From Figure 7.11, estimate the range of possible values for $\Omega_{\Lambda,0}$ at the 99% confidence level. Making the additional assumption that the Universe has a flat spatial geometry, what is the likely range of values for $\Omega_{\Lambda,0}$?

7.4.2 Determination of $\Omega_{m,0}$ from mass-to-light ratios

There are many methods of determining the current value of Ω_m apart from supernova measurements. Some of these methods are of dubious reliability, and recent results for $\Omega_{m,0}$ have ranged from about 0.2 to 1.5, often with substantial uncertainties. A selection of these results is shown in Figure 7.12, where the uncertainty associated with each determination is represented by a horizontal bar. The method used in each of these determinations has been named in the figure, but the details of most of those methods are not explained here. One method, however, is regarded as being particularly important, and that will be discussed in this section. This is the method of mass-to-light ratios.

The **mass-to-light ratio** of an astronomical system, such as a star or galaxy or even a cluster of galaxies, is the value obtained by dividing the mass M of the system by its luminosity L . The luminosity in this definition is normally restricted to some specified range of wavelengths, usually the optical range. So, in the case of the Sun where the mass is M_\odot and the (optical) luminosity is L_\odot , the mass-to-light ratio is given by M_\odot/L_\odot .

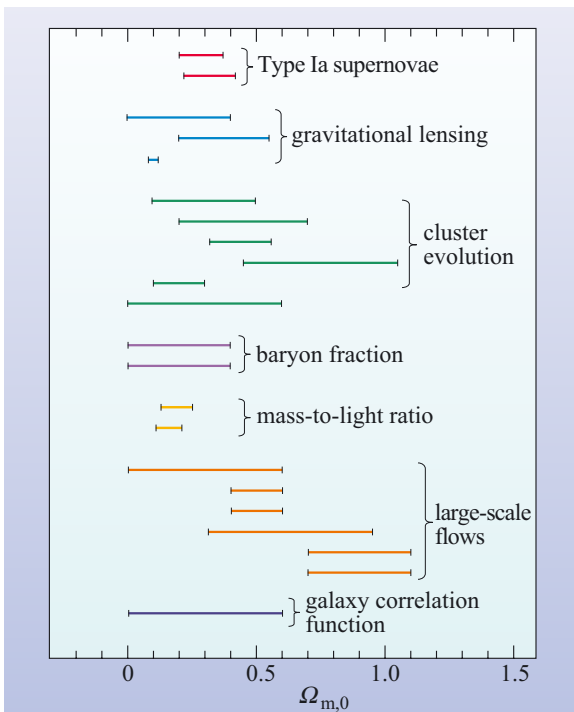


Figure 7.12 A selection of recent results for the current value of Ω_m . In some cases, such as those from supernovae, the actual observational results have been additionally restricted by assuming that the Universe has a flat spatial geometry on the large scale. Observational support for this assumption is provided in Section 7.5. This figure summarizes a great deal of modern research literature and is based on an original figure by Sabine Schindler.

The mass-to-light ratio of the Sun is not of much direct interest in cosmology, but if observers could determine the mass-to-light ratio of a representative portion of the Universe then they could multiply that quantity by the observable luminosity density of the Universe, j_{Univ} , to obtain the average matter density ρ and hence the density parameter for matter Ω_{m} . In terms of symbols

$$\Omega_{\text{m}} = \rho/\rho_{\text{crit}} = [(M/L)_{\text{Univ}} \times j_{\text{Univ}}]/\rho_{\text{crit}} \quad (7.11)$$

where $(M/L)_{\text{Univ}}$ represents the mass-to-light ratio of the Universe.

- What would be appropriate SI units of $(M/L)_{\text{Univ}}$ and j_{Univ} ?
- $(M/L)_{\text{Univ}}$ could be expressed in terms of kg W^{-1} (i.e. kilogram per watt) and j_{Univ} could be measured in W m^{-3} (i.e. watt per cubic metre). In practice, astronomers are quite likely to use M_{\odot}/L_{\odot} as a unit of mass-to-light ratio, and to express the luminosity density in terms of solar luminosities per cubic parsec ($L_{\odot} \text{ pc}^{-3}$), but these are not SI units.

Any attempt to use mass-to-light ratios to determine $\Omega_{\text{m},0}$ must, of course, take account of many details. For instance, the same range of wavelengths must be used in determining $(M/L)_{\text{Univ}}$ and j_{Univ} , and the range over which all quantities are measured, although large enough to be cosmologically representative, must also be small enough to ensure that it is the *current* values of $(M/L)_{\text{Univ}}$ and j_{Univ} that are being determined. (Presumably, the values of both $(M/L)_{\text{Univ}}$ and j_{Univ} have changed over time.) Nonetheless, the basic challenges of this method are clear: determine $(M/L)_{\text{Univ}}$ and measure j_{Univ} .

Over the years, these challenges have been confronted by a number of astronomers, starting with Fritz Zwicky (see Section 4.3.2) in the 1950s. The two attempts to use this method that are shown in Figure 7.12 date from 1997 and 2000 and are, respectively, the work of Raymond G. Carlberg *et al.* and Neta Bahcall *et al.* Bahcall and her collaborators restricted luminosities to blue wavelengths, and it was assumed that $(M/L)_{\text{Univ}}$ would be well represented by the mass-to-light ratio of clusters of galaxies. The masses of the clusters were measured using the optical and X-ray techniques that were described in Chapter 4.

- List and briefly describe the optical and X-ray techniques for determining cluster masses that were described in Chapter 4.
- *The virial mass method*, which is based on Doppler measurements of the velocity dispersion within the cluster and the assumption that the cluster is in dynamical equilibrium (i.e. relaxed or virialized).
- The X-ray emission method*, which is based on X-ray spectra and X-ray surface brightness measurements (to determine the temperature and density distribution of the intracluster gas) and the assumption that the gas is in hydrostatic equilibrium.
- The gravitational lensing method*, which is based on observations of distorted images of more distant galaxies and models of the mass distribution within the lensing cluster.

It has been clear for a long time that mass-to-light ratios of astronomical systems tend to increase as the size of the system becomes larger, at least up to scales of hundreds of kiloparsecs. When dealing with galaxies and clusters of galaxies this

growth is usually interpreted as signifying the presence of large amounts of dark matter, since dark matter tends to increase the mass of a system without producing any corresponding increase in luminosity. In fact, the growth of mass-to-light ratios with the size scale on which they are measured is often quoted as evidence of the existence of dark matter. However, there is also evidence that, as the size of the systems considered increases to a megaparsec and beyond, so the growth in the mass-to-light ratio ceases and M/L attains a constant value. It is this constant limiting value that is taken to represent the average ‘cosmic’ value of the mass-to-light ratio.

Figure 7.13 provides a highly schematic representation of the way that observed mass-to-light ratios increase with the scale on which they are measured. As you can see, for the visible parts of galaxies (corresponding to measurement scales of about 10 kpc) the mass-to-light ratio is about five times M_\odot/L_\odot , while on the larger scale of entire galaxies (a scale of 100 kpc or so, that would include the galaxy’s dark halo) the mass-to-light ratio is about 20 times M_\odot/L_\odot , and on the even larger scale of groups and clusters (scales around 1 Mpc) the mass-to-light ratio is more like 240 times M_\odot/L_\odot . The belief that the mass-to-light ratio becomes roughly constant on the largest of these scales is supported by various measurements, including the finding that the mass-to-light ratio of the supercluster MS302 is essentially equal to the mass-to-light ratios of the clusters of which it is composed. (In this latter case, all the masses were determined using the gravitational lensing method.)

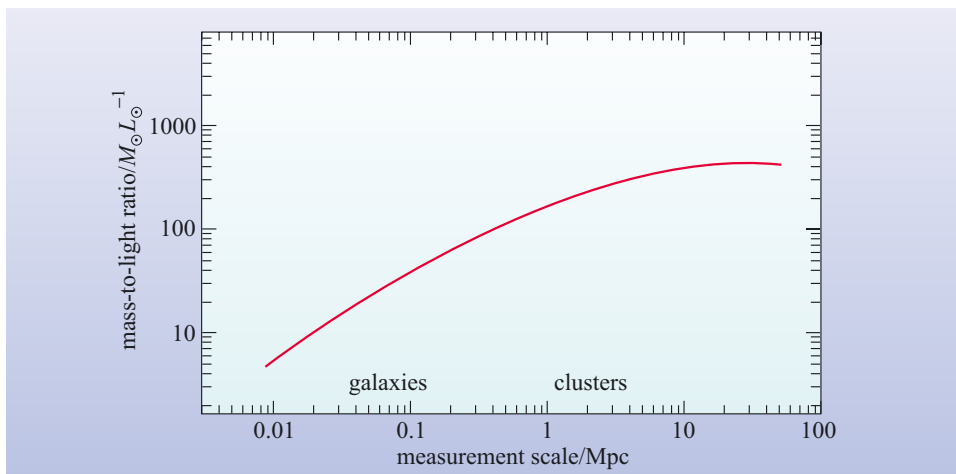


Figure 7.13 A schematic representation of the mass-to-light ratio as a function of measurement scale.

According to Bahcall *et al.*, mass and light do share the same overall distribution on the large scale (i.e. the mass-to-light ratio does become constant) but there is still a great need for caution when making measurements because relatively dense regions such as rich clusters of galaxies exhibit higher mass-to-light ratios than do lower density regions. This so-called ‘bias’ effect is attributed to the relatively more highly evolved state of the high-density regions, where the emission of blue light from stars has declined more than in the relatively less evolved, lower density regions. Taking this bias into account, their conclusion is that

$$\Omega_{m,0} = 0.16 \pm 0.05 \quad (7.12)$$

This is somewhat lower than the widely favoured value obtained from supernova cosmology ($\Omega_{m,0} \approx 0.3$), but not vastly different given the difficulty of the measurements and the range of results indicated in Figure 7.12. Some cosmologists regard the estimated uncertainty in Equation 7.12 as rather optimistic. Perhaps the

real significance of these results is the support they give to the idea of dark matter becoming increasingly important on larger size scales, and the insight they provide into the poorly understood relationship between the distribution of light and mass.

7.4.3 Determinations of $\Omega_{b,0}$

The methods of density determination that have been discussed in the last two sections have both been sensitive to dark matter, and have indicated its presence in substantial amounts. On the basis of mass-to-light ratios, it is widely assumed that much of this dark matter is contained in dark-matter halos of galaxies. However, these measurements give little indication of the nature of the dark matter. One obvious possibility is that it might be some form of ordinary baryonic matter. One way to investigate this possibility is to determine the density parameter for baryonic matter $\Omega_b(t)$; if the current value of this quantity, $\Omega_{b,0}$, is significantly less than the current value of the density parameter for all kinds of matter, $\Omega_{m,0}$, then much of the dark matter must be non-baryonic.

This section presents two arguments for the belief that the current average density of baryonic matter is much less than the current total density of matter. The first of these arguments is based on the requirement that the relative abundances of light elements predicted by the theoretical account of *primordial nucleosynthesis* (as discussed in Chapter 6) should agree with the observational evidence regarding those abundances. The second argument is based on direct observational assessments of the density of baryonic matter in various parts of the Universe; that is to say a *baryon inventory*. In addition to these two arguments, a third argument based on observations of the cosmic background radiation is presented in Section 7.5.

Primordial nucleosynthesis

According to Chapter 6, the abundances of light elements (essentially hydrogen, helium and lithium) predicted by primordial nucleosynthesis are related to the current value of the density parameter for baryonic matter. This link exists because the total number of baryons in the Universe is a *conserved* quantity that is not expected to change with time. So the current mean density of baryonic matter is related to the density of baryonic matter at the time of cosmic nucleosynthesis, and this primordial density played an important role in determining the extent to which light nuclei were synthesized between about three and 30 minutes after the start of cosmic expansion.

One of the great successes of big bang cosmology is its ability to simultaneously predict primordial abundances of deuterium, helium and lithium that are consistent with the abundances currently observed in stars and gas clouds (once allowances have been made for more recent astronomical processes such as stellar nucleosynthesis). In order to achieve this consistency, however, it is necessary that the current value of the density parameter for baryonic matter should be in the range

$$0.02 \leq \Omega_{b,0} \leq 0.05 \quad (7.13)$$

This is much lower than the estimates of $\Omega_{m,0}$ based on supernova cosmology which, as you saw earlier, are around 0.3. So it seems that baryons account, very roughly, for only somewhere between about a fifteenth and a sixth of the total density of matter in the Universe.

A baryon inventory

Where are the baryons (protons, neutrons, etc.) in the Universe? Well, some are in you and me. We each contain about 2 to 4×10^{28} protons and neutrons. But there

are many more in the Earth, far more in the Sun, and enormously more in the Milky Way and other galaxies. Perhaps surprisingly, stars and their remnants are thought to account for less than 30% of the baryonic matter in the Universe. The majority of the baryons are believed to reside in the various forms of ionized gas that exist within and between the galaxies.

Table 7.2 is adapted from a baryon inventory that was published in 1998. The table shows the estimated contribution to $\Omega_{b,0}$ from various sources. The authors of the inventory based their estimates on a wide range of observational data. When dealing with the well-determined contribution from hot, ionized gas in rich clusters they made use of the kind of X-ray observations described in Chapter 4. X-ray observations at somewhat lower energies provided estimates of the cooler ionized gas in sparser groups of galaxies, while Lyman α absorption measurements (also described in Chapter 4) allow the quantity of ionized gas in cool clouds to be inferred. Unfortunately these last two estimates, although important to the final sum, are less reliable than many of the other values that appear in the table.

Table 7.2 The main repositories of baryons in the Universe, and their respective contributions to the current value of the density constant for baryonic matter.

Repositories of baryons	Contribution to $\Omega_{b,0}$
stars in elliptical galaxies and the spheroids of spirals	0.0026
stars in discs (for spiral galaxies)	0.0009
stars in irregular galaxies	0.0001
neutral atomic gas	0.0003
molecular gas	0.0003
ionized gas in clusters of galaxies	0.0026
ionized gas in groups of galaxies	0.0056
ionized gas in cool clouds	0.0020

A baryon inventory of this kind is open to the criticism that many of the entries, including some of the most important, are hard to estimate. It is also the case that some repositories of baryons may have been overlooked. In Table 7.2, for example, it is certainly the case that warm, ionized gas in the voids between superclusters has been ignored, and so has the contribution of baryons in dwarf galaxies and in galaxies of low surface brightness. The contribution from these and other neglected sources are thought to be small, but there is inevitably some uncertainty about this.

The general conclusion, after taking account of all known repositories of baryons and making reasonable estimates wherever data are lacking, is that $\Omega_{b,0}$ is in the range

$$0.007 \leq \Omega_{b,0} \leq 0.041 \quad (7.14)$$

with a ‘best guess’ value of

$$\Omega_{b,0} \approx 0.021 \quad (7.15)$$

Because of the uncertainties this ‘best guess’ should not be taken too seriously. However, the upper part of the range given in Equation 7.14 is consistent with the requirements of primordial nucleosynthesis, and gives further evidence that the dark matter cannot be entirely, or even mainly, baryonic matter.

QUESTION 7.5

Justify the rough claim made earlier that ‘baryons only account for somewhere between about a fifteenth and a sixth of the total density of matter in the Universe’.

7.5 Anisotropies in the cosmic microwave background and precision cosmology

The cosmic microwave background (CMB) radiation was introduced in Chapter 5 and then explored more fully in Chapter 6. Among the many features of the CMB that were described in those earlier discussions were the following.

- 1 The CMB radiation is intrinsically uniform to better than one part in 10 000. This means that, after correcting for effects due to the motion of the detector, the CMB comes with very nearly equal intensity from all directions.
- 2 The CMB is ‘thermal radiation’ with a characteristic temperature of 2.725 K. This means that the intrinsic spectrum of the radiation is well described by a black-body curve with a peak in intensity at a wavelength of about 1 mm.
- 3 The CMB radiation is believed to have originated at the time of recombination, when the mean temperature of cosmic matter became low enough to allow electrons and protons to form hydrogen atoms that were not immediately ionized again. This means that the photons that make up the CMB were last scattered about 3 to 4×10^5 years after the start of cosmic expansion, when the temperature was about 3000 K.
- 4 The radiation has been expanding and cooling since it was last scattered, during which time its temperature has decreased by a factor of about 1100. This means that the ‘last-scattering surface’ from which the radiation was released is now at a redshift of about 1100. (The most distant observed galaxies are at a redshift of about 6.)
- 5 Despite its high degree of intrinsic uniformity, anisotropies in the intensity of the radiation have been observed, at a level of a few parts in 100 000, over a range of angular scales. These tiny variations in intensity represent fluctuations in the temperature of the CMB and are believed to hold the key to the precise measurement of the cosmological parameters.

This section is concerned with the anisotropies in the CMB, and discusses the origin of the anisotropies, how they can be measured and described, and how they can be used to determine the key cosmological parameters. It also contains very recent results that improve and make more precise some of the values quoted in earlier chapters and in the points listed above.

QUESTION 7.6

In the numbered list above, no mention is made of the ‘dipole anisotropy’ that was introduced in Chapter 6. Briefly describe the dipole anisotropy, explain its origin and identify the words and phrases in points 1, 2 and 5 that imply the dipole anisotropy has been taken into account, despite the lack of any explicit reference to it.

7.5.1 Detecting the anisotropies in the CMB

The discovery of the CMB by Penzias and Wilson in 1965 was a turning point in observational cosmology. Following the discovery, many observers reoriented their research in order to concentrate on the CMB and the other implications of big bang cosmology. Penzias and Wilson had made their serendipitous discovery using a ground-based radio detector working at a wavelength of about 70 mm, but much of the subsequent work tried to cover wavelengths closer to the expected peak in the CMB spectrum, at about 1 mm. At such relatively short wavelengths, radiation coming from space is absorbed by the Earth's atmosphere, so many of the observations were made using detectors carried by balloons or high-flying aircraft. The existence of the dipole anisotropy was first established in this way, in 1977, using a detector fitted to an American U-2 aircraft. (U-2s had been employed as spy planes during the Cold War, but were extensively used for scientific work from the mid-1970s.) Other high-altitude work gave a strong indication that the CMB spectrum was thermal (i.e. described by a black-body curve), but left open the possibility that there might be small departures from a pure black-body spectrum. However, none of the balloon or aircraft observations revealed any anisotropies in the CMB, apart from the large-scale dipole anisotropy. Small-scale anisotropies were predicted by big bang cosmology, but highly sensitive detectors with at least moderate resolution were needed to identify them. The observation of those small-scale anisotropies, and the confirmation of the purely thermal nature of the CMB spectrum, had to await the launch of the first space satellite to be dedicated to the study of the cosmic background radiation.

COBE, the Cosmic Background Explorer (Figure 7.14) was launched on 18 November 1989. COBE was about the size and mass of a large car, and carried three main experimental packages:

- DMR, a set of three differential microwave radiometers designed to search for anisotropies in the cosmic background radiation;
- FIRAS, the Far-Infrared Absolute Spectrophotometer, designed to measure the CMB spectrum; and
- DIRBE, the Diffuse Infrared Background Experiment, designed to search for cosmic infrared background radiation, and to assess local sources of diffuse infrared radiation, including the Milky Way.

Within a short time of starting its astronomical observations, the FIRAS detector gave powerful support to the big bang prediction of a very pure black-body spectrum. It is now known that, over the wavelength range from 0.1 mm to 5 mm, the CMB spectrum follows a 2.725 K black-body spectrum very precisely. Observations of small-scale anisotropies by the DMR took longer to emerge and longer still to be refined.

As mentioned in Chapter 6, absolute measurements of the intensity of the CMB are difficult to make, so the DMR was designed to look for differences in the intensity of the CMB from pairs of regions separated by 60° on the sky.



Figure 7.14 COBE, the Cosmic Background Explorer. COBE's three main scientific instruments, FIRAS, DMR and DIRBE, made their observations from behind a collapsible thermal shield that was deployed after the satellite was placed in orbit. COBE followed a relatively low Earth orbit that took it over the Earth's poles. The plane of the orbit precessed at the rate of about one degree per day; this allowed the shielded instruments to scan all parts of the sky over a period of six months. (NASA)

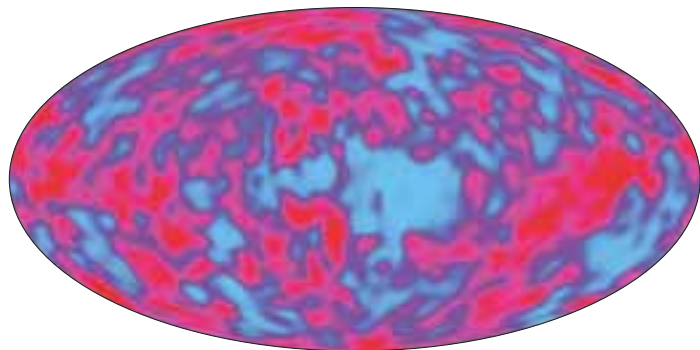


Figure 7.15 An all-sky CMB anisotropy map based on data from the DMR experiment on COBE. The angular resolution of this map is about 7° . Contrast this map with the highly isotropic distribution measured at lower sensitivities, shown in Figure 5.4. (Edward L. Wright)

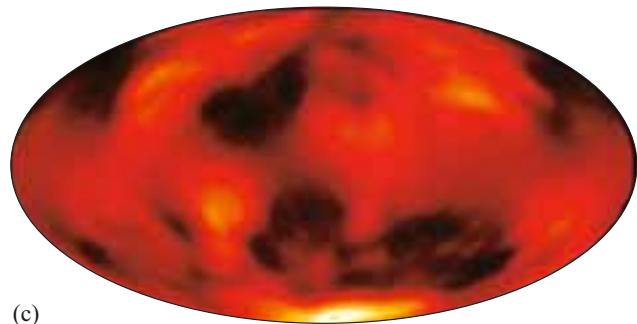
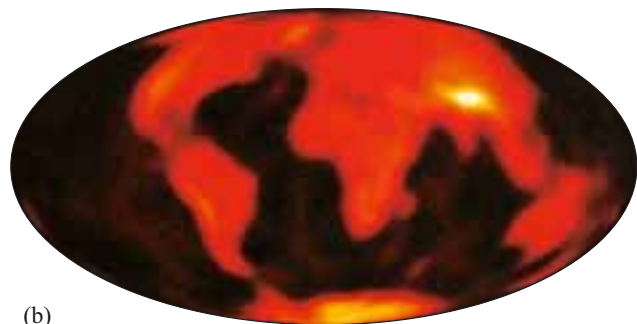
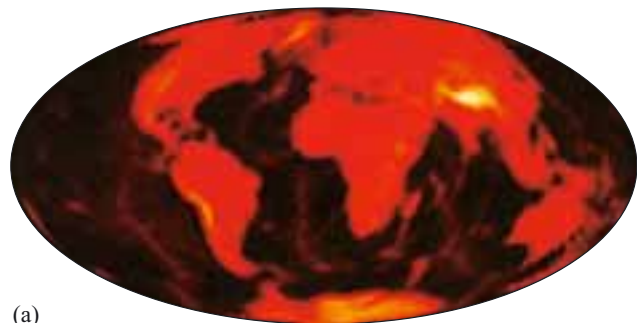


Figure 7.16 (a) A map of the Earth. (b) The same map viewed with the angular resolution that is inherent in the COBE anisotropy map. (c) The addition of noise makes the underlying pattern even harder to discern. (Courtesy of Ted Bunn)

The same technique had been used to detect the dipole anisotropy in the 1970s, and the DMR on COBE was a direct descendant of the detector that had been carried to high altitude by a U-2 spy plane. The COBE DMR actually consisted of three pairs of detectors working at wavelengths of 3.3, 5.7 and 9.6 mm. By comparing observations at these three wavelengths, the effects of microwave emission from the Milky Way could be identified and eliminated, leaving just the cosmic signal, albeit contaminated by unavoidable detector ‘noise’. Thanks to COBE’s rotation (it turned on its axis once every 75 seconds), the satellite’s 103-minute orbital period, and the gradual precession of the orbital plane, the DMR was eventually able to examine all parts of the sky and to produce the all-sky CMB anisotropy map shown in Figure 7.15.

The publication of the first anisotropy map in April 1992 was another very significant development in observational cosmology. Nonetheless, it is important to recognize the limitations of the COBE findings. In the first place, the data on which the map was based were very noisy. This means that much of the visible structure in the map may be illusory. A statistical analysis of the data justified the claim that anisotropies had been observed, but it did not guarantee that the coloured boundaries that appear in the map were accurately located. Moreover, the nature of the DMR’s collecting horns meant that each of the paired regions being compared was about 7° across, so this was the minimum angular scale on which anisotropies might reliably be detected. To give some idea of the significance of this, Figure 7.16 shows a map of the Earth viewed with the same kind of resolution. As you can see, even the major continents are indistinct, while finer details, such as the existence of the British Isles, are completely lost. The features become even more obscure when ‘noise’ is deliberately added to the data to make the comparison with the DMR results more apposite. Clearly, despite COBE’s success, much work remained to be done.

In the early 1990s, further measurements and analyses by the COBE Science Team confirmed and refined the original results for angular scales of about 7° . Since then other groups of researchers, mainly using ground-based or balloon-borne equipment, have struggled to measure CMB anisotropies on smaller angular scales. One particularly notable effort was that of the BOOMERanG collaboration, an international team that used equipment suspended from a high-altitude balloon (Figure 7.17). BOOMERanG (which stands for ‘balloon observations



Figure 7.17 The high-altitude balloon and instrument package used by the BOOMERanG team to measure CMB anisotropies from Antarctica. (BOOMERanG Collaboration)

of millimetric extragalactic radiation and geophysics’) mapped CMB anisotropies over only a limited region of sky, but did so with high sensitivity and an angular resolution of better than 1° . The greatly improved resolution was achieved by cooling the detectors to 0.28 K and placing them at the focus of a small (1.3 m aperture) telescope. The BOOMERanG team published their results in Spring 2000. They were based on 10.5 days of continuous observation from an altitude of 35 km above Antarctica. The Antarctic air is exceptionally clear, clean and dry, so the effects of absorption are minimized there. Some of the results from BOOMERanG are shown in Figure 7.18.

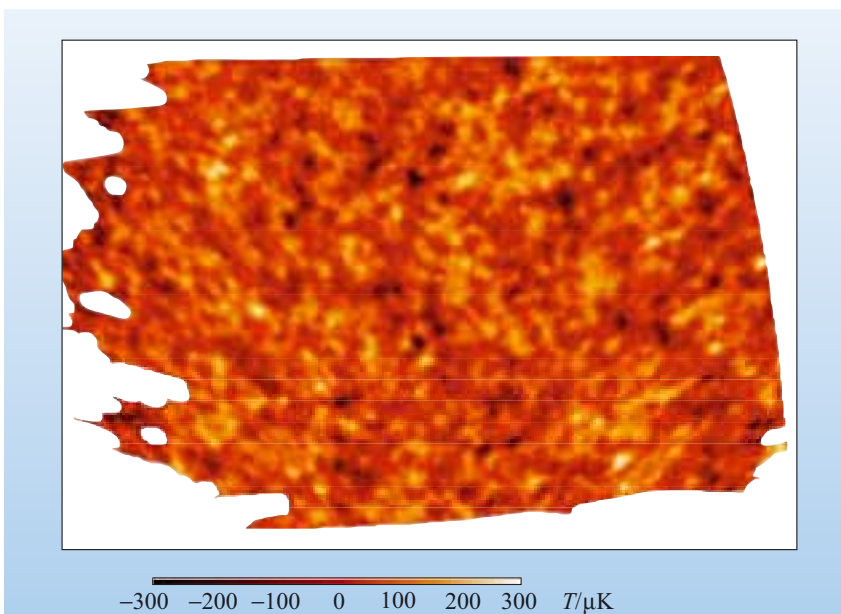


Figure 7.18 Some results from BOOMERanG, which examined only a limited part of the sky but did so at an angular resolution of about 1° . The CMB anisotropies are portrayed as variations in the effective temperature of the radiation. Note that the range of variation is only $\pm 300 \mu\text{K}$, while the mean temperature is about 3 K. (BOOMERanG Collaboration)

Figure 7.19 shows a comparison between the CMB anisotropies actually observed by BOOMERanG and those predicted by big bang cosmology in the context of some Friedmann–Robertson–Walker models with different values of the curvature parameter k . As you can see, even a simple comparison ‘by eye’ favours the model with $k = 0$. This conclusion is supported by more detailed analyses of the data that will be outlined later.

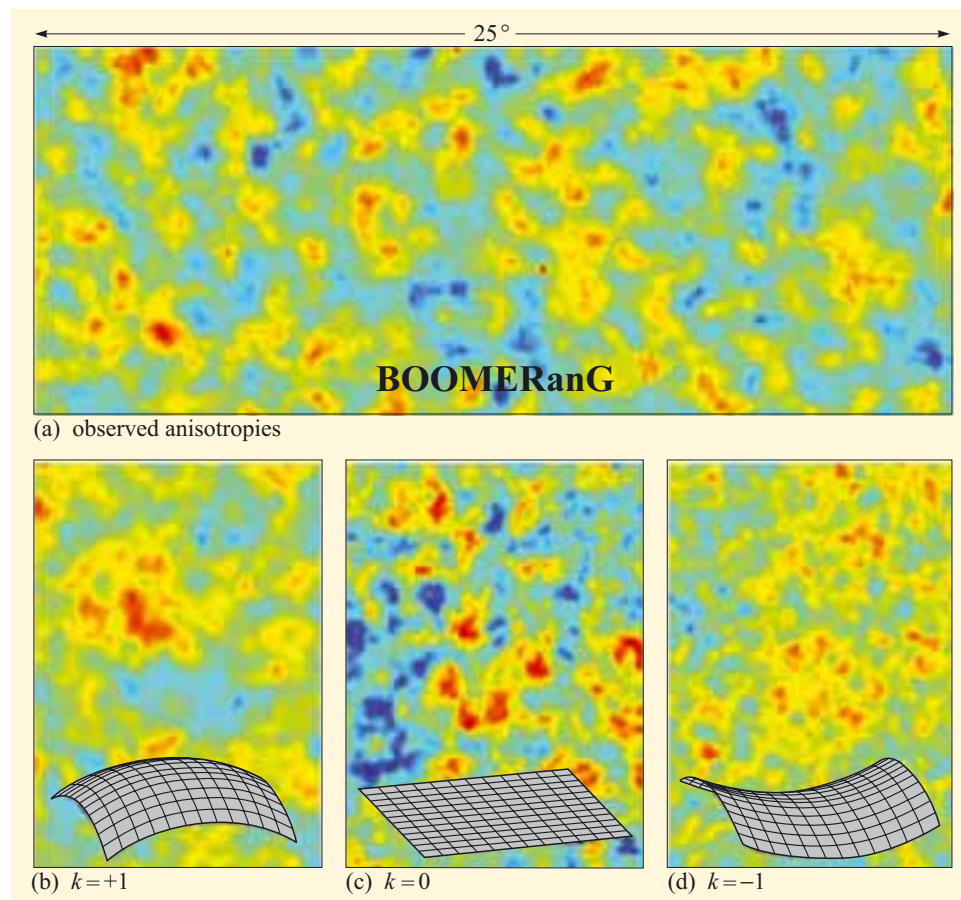


Figure 7.19 Evidence that the Universe has a flat ($k = 0$) spatial geometry. (a) When the anisotropies observed by BOOMERanG are compared with computer predictions (b, c and d) based on various cosmological models, it is the $k = 0$ predictions that provide the best agreement with the observations. The small grids printed at the bottom of the computer simulations provide a reminder of the geometric properties that correspond to the different values of k . (BOOMERanG Collaboration)

Many other degree-scale anisotropy observations have followed those of BOOMERanG, some from balloon experiments, others from ground-based detectors. However, the most recent results at the time of writing are the keenly awaited findings of an American space mission, the Wilkinson Microwave Anisotropy Probe (WMAP), which was launched in June 2001 (Figure 7.20). This is the successor to COBE and, like COBE, it has produced an all-sky anisotropy map, although this time with an angular resolution of about 0.1° . The WMAP anisotropy map is shown in Figure 7.21; it is based on the first year’s observations and may be refined somewhat by further observation and analysis. However, as you can see, it is already far more detailed than the COBE map that was produced roughly ten years earlier.

Analysis of the small-scale anisotropies that can be seen in Figure 7.21 has already resulted in the publication of revised values for a whole range of cosmological parameters, with many of them being quoted to unprecedented levels of precision. How these values have been derived is explained later; first we must examine how the observed anisotropies can be described mathematically, so that the analysis can begin.

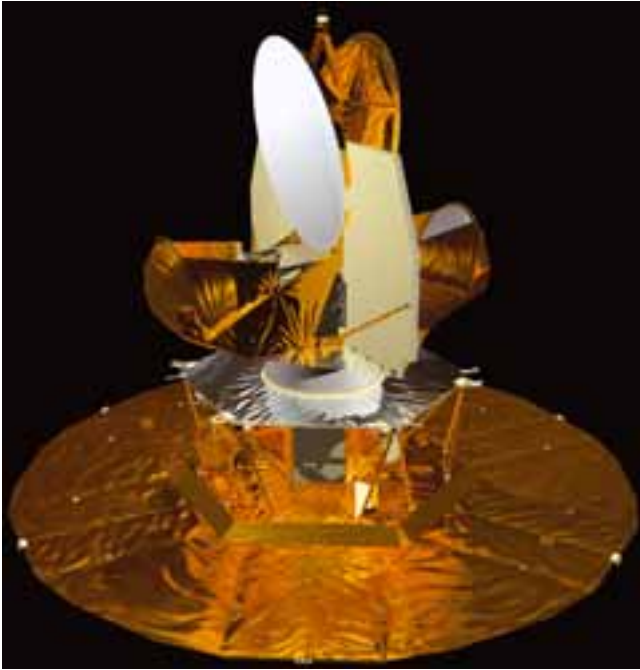


Figure 7.20 Computer image of the WMAP spacecraft. (NASA)

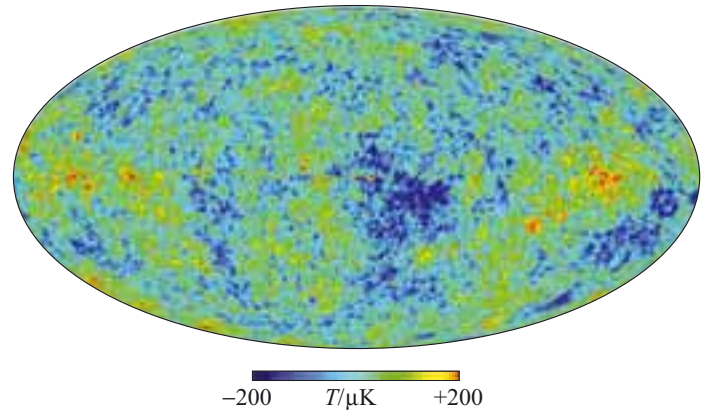


Figure 7.21 An all-sky CMB anisotropy map, based on data obtained by the WMAP space probe. The angular resolution of this map is about 0.1° . Compare this map with the lower resolution COBE data shown in Figure 7.15. (Bennett *et al.*, 2003)

QUESTION 7.7

When comparing the predictions with the observations in Figure 7.19, would you expect the big bang prediction to show precisely the same pattern of anisotropies as that observed by BOOMERanG, or simply something ‘similar’ to the observations? Justify your answer.

7.5.2 Describing the anisotropies in the CMB

Any observation of anisotropies in the CMB (such as shown in Figure 7.21), made from the Earth or anywhere in the Solar System, represents nothing more than a single ‘sample’ of a cosmic phenomenon. The CMB, viewed from anywhere else in the Universe at the present time, is expected to be similar in its general aspects but different in detail. From a cosmologist’s point of view, what is interesting about any single view is not the detailed information it provides about the directions in which the CMB is slightly warmer or slightly cooler, but the *statistical* insight it provides into CMB anisotropies in general. This is analogous to saying that cosmologists value observations of the large-scale distribution of galaxies (described in Chapter 4) not for what they reveal about the locations of particular clusters of galaxies, but rather for the statistical information they provide about the scale of clusters in general. The locations are a *local* concern, the scales are a *cosmic* one.

The quantity that contains all the *statistically* important data in an anisotropy map such as Figure 7.21 is called the **angular power spectrum**. It is constructed from the same data that are used to plot the map, but it effectively discards the detail that depends on our particular location, and makes apparent the cosmically important features of the data. It provides a way of showing graphically the strength of

fluctuations on different scales and bears comparison with methods used to describe large-scale structure in Chapter 4. The angular power spectrum derived from the first year of WMAP observations is shown in Figure 7.22. The aim of this section is to introduce the main features of such a spectrum, and to explore the relationship between an anisotropy map and the corresponding angular power spectrum.

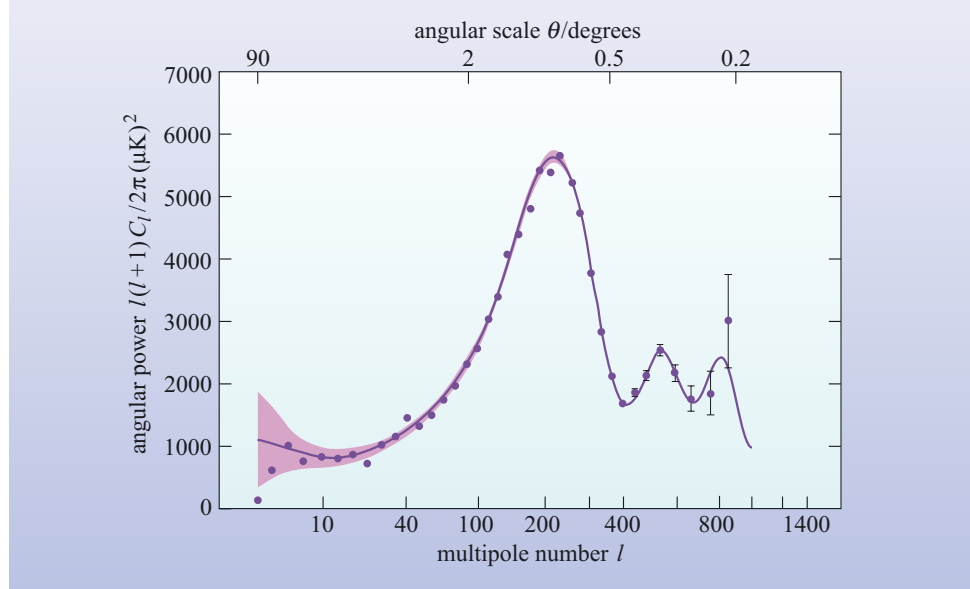


Figure 7.22 The angular power spectrum of the CMB as determined by WMAP. (Bennett *et al.*, 2003)

The data points and uncertainty ranges plotted in Figure 7.22 show what WMAP has revealed about the angular power spectrum of the CMB as a whole. The smooth curve represents the ‘best fit’ to these data from a range of theoretical predictions based on various FRW models (more will be said about these predictions in the next section), and the shaded band enclosing the best-fit line represents an effect known as **cosmic variance**. The cosmic variance is a consequence of the fact that we are compelled to estimate the angular power spectrum of the CMB from observations made at one cosmic location (i.e. on the basis of a single ‘sample’ of the CMB, made from within the Solar System). As a result, even though theories may predict a precise form for the CMB’s angular power spectrum (represented by the best-fit line), measurements of that spectrum from a single location should only be expected to show consistency with the broadened band that surrounds the line.

The angular power spectrum is usually presented as a graph in which a variable called *angular power* is plotted along the vertical axis, while an independent variable called *multipole number* is plotted along the horizontal axis. Both of these variables are explained below.

In the context of CMB anisotropy studies, the **multipole number** (usually denoted by the italic letter l), provides a natural, if somewhat unfamiliar, way of describing angular separations. The difference between two directions can always be described in terms of the angle θ between them. The angle θ may be associated with a multipole number l using the relation

$$l = 180^\circ/\theta \quad (7.16)$$

implying that the multipole number l is the number of times the angle θ can be fitted into a 180° arc. Because of this relationship between multipole number and angle, the horizontal axis of an angular power spectrum is sometimes shown as an angular

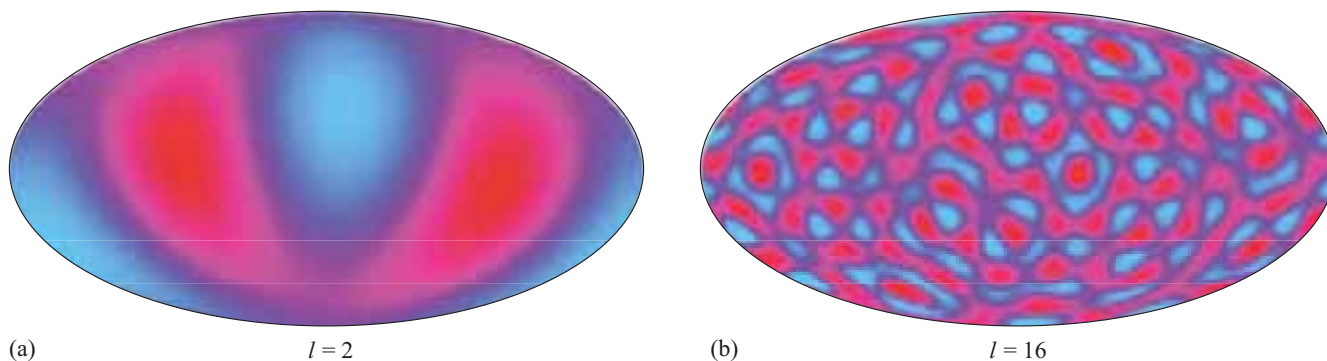
scale, but with the angle *decreasing* in unequal steps from left to right. (A scale of this kind is shown at the top of Figure 7.22.) In what follows we mainly make use of the multipole number, but we sometimes refer to the corresponding angle derived from Equation 7.16.

- COBE was only able to detect anisotropies of angular size 7° or above. What is the corresponding range of l values?
- From Equation 7.16, $\theta = 7^\circ$ corresponds to an l value of about 26. Larger angles correspond to smaller values of l , so we can say that the COBE results were confined to the range $l = 0$ to $l = 26$ or thereabouts. Looking at Figure 7.22 you can see that this is a very small range compared with the WMAP results; it doesn't even cover the first peak.

For any given value of l in Figure 7.22, the corresponding value of the **angular power**, usually denoted $l(l+1)C_l/(2\pi)$, indicates how much variation is present in the anisotropy measurements on that angular scale. The precise definition of angular power is complicated and too technical to merit discussion here. However, the important quantity C_l can be determined directly from the data used to plot the anisotropy map, and reveals the extent to which the temperature measured at one point will, on average, be *correlated* with the temperature at some other point an angular distance θ away. To get some feeling for the significance of this, consider the two artificial anisotropy maps shown in Figure 7.23. The map in Figure 7.23a shows no small-scale variation in the temperature; the anisotropy is on a scale of 90° , and the corresponding angular power spectrum would show a peak at $l = 2$. In contrast, the map in Figure 7.23b consists of anisotropies that share a common angular scale of about 11° ; in this case the corresponding angular power spectrum would peak at $l = 16$. In neither case would the angular power spectrum reveal anything about the precise location of the coloured boundaries in the maps, but it would indicate the angle that, on average, separates a region of one colour from the next region of that colour.

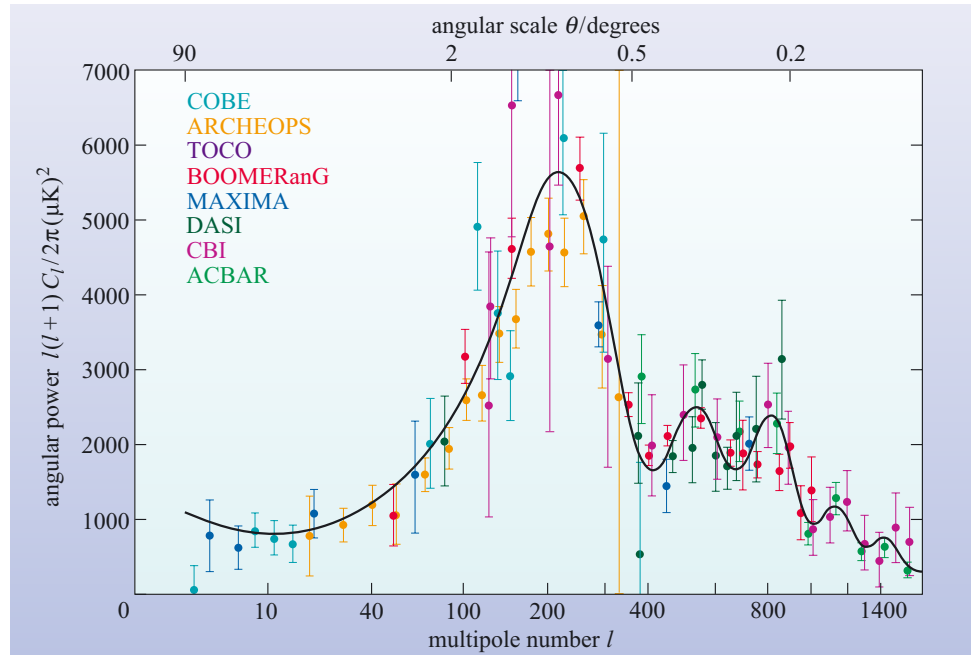
The WMAP results represent a major contribution to the effort to determine the angular power spectrum of the CMB anisotropies. Over the range of l values shown, the WMAP measurements provide almost as much information about the angular power spectrum as it is possible to obtain from a single location in the cosmos. Figure 7.24 shows a compilation of recent measurements (excluding those from WMAP), along with the best-fit line to the WMAP data shown in Figure 7.22. As you can see, despite some large uncertainties in the older measurements, there is an impressive level of agreement about the shape of the

Figure 7.23 Two artificial all-sky maps showing different patterns of anisotropy. (a) A large-scale anisotropy for which all the angular power is concentrated at $l = 2$. (b) Smaller anisotropies that share a common angular scale of about 11° , and for which the angular power is concentrated at $l = 16$. (Edward L. Wright)



angular power spectrum. As Figure 7.24 indicates, the observed CMB anisotropies have some angular power at all angular scales, but there is a pronounced peak around $l = 220$, which corresponds to an angle of slightly less than 1° , and there are lesser peaks at larger values of l . These peaks show the angular scales at which temperature variations in the CMB are strongest. The next section concerns our theoretical understanding of those peaks in the angular power spectrum, and our ability to predict them on the basis of big bang cosmology.

Figure 7.24 A compilation of recent measurements of the angular power spectrum of CMB anisotropies (excluding those from WMAP), together with the best-fit line obtained from the WMAP data. There is good agreement, despite some large uncertainties in some of the measurements. (Hinshaw *et al.*, 2003)



QUESTION 7.8

At what values of l are the second and third peaks in the angular power spectrum in Figure 7.24? What are the corresponding angular scales? What is the maximum angular power of these two peaks according to Figure 7.24?

7.5.3 Predicting the anisotropies in the CMB

Cosmologists treat the Universe as highly uniform on the large scale, but they know that this is not a valid approximation on the small scale. The existence of people, planets, stars, galaxies and clusters of galaxies all demonstrate that there are non-uniformities in the Universe, and the anisotropies in the CMB provide our earliest view of those non-uniformities. An all-sky CMB anisotropy map, such as Figure 7.21, is essentially a snapshot of the spherical surface, centred on the Earth, at which the CMB radiation now reaching us was last scattered. That surface was discussed in Chapter 6, where it was referred to as the *last-scattering surface*. According to Chapter 6, the last scattering occurred at the time of recombination, when matter and radiation became decoupled. The time of this decoupling is usually supposed to be about 3 to 4×10^5 years after the start of cosmic expansion, but the recent WMAP data support a more precise value: $t_{\text{dec}} = 3.8 \times 10^5$ years, which we adopt from here on. So Figure 7.21 shows that when the Universe was about 380 000 years old some parts of it were slightly warmer or slightly cooler than others.

The existence of these **temperature fluctuations** at the time of decoupling indicates the presence of a corresponding set of **density fluctuations** in the distribution of cosmic matter (i.e. regions where the density of matter was slightly higher or slightly lower than the average density). Chapter 6 explained the link that is thought to exist between density fluctuations in the early Universe and the large-scale clustering of galaxies that we see in the present-day Universe. In this section, our main concern is the relationship between the density fluctuations and the anisotropies in the CMB. In particular we shall see how the effects of density fluctuations can be related to the features seen in the CMB's angular power spectrum. Note that the emphasis here is on *statistical* features. You will shortly see how the density fluctuations can be characterized statistically, and how that statistical characterization allows us to predict the *statistical* features of the CMB anisotropies. No attempt is made to predict the particular pattern of anisotropies shown in the all-sky map of Figure 7.21: it is the statistically significant angular power spectrum of Figure 7.22 that is of interest.

It is widely assumed that the density fluctuations present in the early Universe first arose as a natural consequence of quantum processes in the very early Universe, and were then amplified by the process of *inflation* that was discussed in Chapter 6. One of the exciting things about modern observational cosmology is that it is now beginning to test this assumption, and some speculative proposals concerning inflation have already been ruled out by measurements of the CMB. However, whatever the role of inflation may have been, there is general agreement that the early Universe contained density fluctuations on essentially all size scales and with a range of strengths. Any detailed assumptions we make about the relative strengths of fluctuations of different sizes can be presented in the form of an assumed 'spectrum' of fluctuations, and this can be characterized mathematically by introducing some more cosmological parameters. The simplest credible spectrum of density fluctuations involves two new cosmological parameters: a *power spectrum normalization*, denoted A , and a *scalar spectral index*, denoted n_s . We shall not be much concerned with these two parameters, but we should add them to the list of cosmological parameters that observational cosmologists are trying to determine, and keep in mind the possibility that more parameters might be needed to fully describe the fluctuation spectrum.

To work out how a spectrum of density fluctuations, characterized by A and n_s , influences the anisotropy of the CMB, we need to consider the interaction of matter and radiation in the early Universe, at least up to the time of decoupling ($t_{\text{dec}} = 3.8 \times 10^5$ years). This is best done by considering the various constituents of the Universe separately, and then considering the way they interact with each other.

Before decoupling, most of the matter in the Universe would have been non-baryonic dark matter, just as it is now, since the relative densities of baryonic and non-baryonic matter are not expected to change with time. The density fluctuations would have been largely composed of dark matter, and can be roughly thought of as dark matter 'halos', of various densities and sizes. These halos would have been expanding, like the rest of the Universe, but those that had a slightly greater than average density would have been growing a little less rapidly than average, and would have been increasing in 'strength' as a result.

The radiation in the Universe, prior to decoupling, would have been plentiful and energetic. As in the present-day Universe, the radiation would not have had much direct interaction with the dark matter. However, the radiation would have incessantly interacted with the electrically charged particles of baryonic matter (such as nuclei of hydrogen and helium). As a result, the baryons and the radiation would jointly form a

sort of fluid, with a common temperature and pressure. This **photon–baryon fluid** would be subject to the gravitational influence of the dark-matter halos and would be drawn towards those halos. However, as indicated in Chapter 6, the gravitational tendency to compress the photon–baryon fluid would have been resisted by the internal pressure of the fluid.

When the Universe had been expanding for a time t , effects due to the pressure in the photon–baryon fluid would not have been able to make themselves felt over distances greater than ct . (The speed of light represents an upper limit to the speed at which signals of any kind can travel.) Consequently, the photon–baryon fluid inside a dark-matter halo that is larger than ct at time t expands along with the halo, essentially undisturbed by the effects of pressure. However, as time passes, t increases and the scale of pressure effects grows. As ct comes to exceed the size of any particular dark-matter halo, the photon–baryon fluid collapsing into that halo will be subject to an increasing pressure that will halt the collapse, and then allow the photon–baryon fluid to ‘spring back’. The photon–baryon fluid contained in some dark-matter halos can undergo several of these oscillations, allowing the density fluctuations to act as the generators of **acoustic waves** (i.e. sound waves) in the photon–baryon fluid. An important point to remember is that at time t the greatest wavelength that any of these waves may have will be roughly ct . We make use of this later.

When the Universe gets to be about 380 000 years old, recombination occurs. This allows the photons to decouple from the baryons. It is the decoupled photons streaming away from the surface of last scattering at $t_{\text{dec}} \approx 380\,000$ years that eventually produce the CMB we now observe. The anisotropies that we see in the CMB are mainly caused by the unevenness of the last-scattering surface, and this is due to the density fluctuations and the acoustic waves that are present there. Now that we have identified the cause of the anisotropies we can account for the main features of their angular power spectrum.

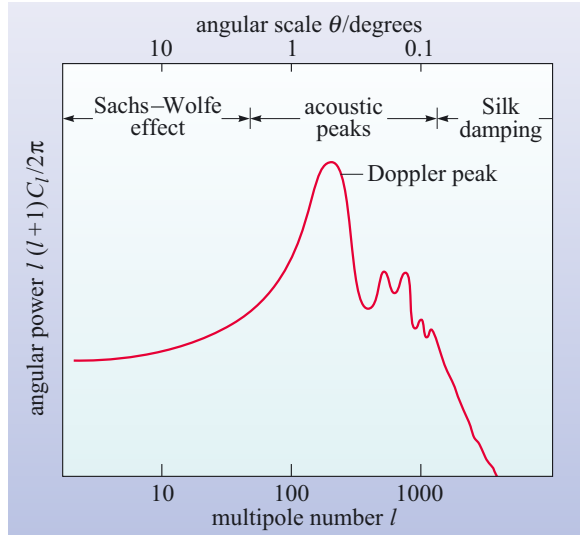


Figure 7.25 The causes of the main features in the angular power spectrum. The various effects named on the graph are discussed in the text. Note that in this case the multipole number has been plotted logarithmically, to allow the inclusion of large values of l .

On angular scales of a few degrees or more (i.e. multipole numbers of about 50 or less), the main source of anisotropy is called the **Sachs–Wolfe effect**. This is largely due to the general relativistic phenomenon of *gravitational red-shift*, which causes photons coming from the denser parts of the last-scattering surface to be observed with slightly longer wavelengths than identical photons leaving other parts of the last scattering surface. This causes the denser parts to appear slightly cooler and leads to the flat plateau seen on the left of the angular power spectrum in Figure 7.25.

On intermediate angular scales (i.e. multipole numbers between 50 and 1000, say), the angular power spectrum shows the effect of the acoustic waves at the time of decoupling. At that time there are some long wavelength waves that are just reaching their state of maximum compression for the first time. These compressions heat the photon–baryon fluid, causing the photons that escape from them to create temperature anisotropies. The moving charged particles associated with the waves will also cause the wavelengths of scattered photons to change (this is a Doppler effect), and this too will create temperature anisotropies. The size scale of the anisotropies would be comparable to the wavelength of the waves, and of the order of ct_{dec} .

- Evaluate ct_{dec} in light-years.
- Since $t_{\text{dec}} \approx 380\,000$ years, it follows that $ct_{\text{dec}} \approx 380\,000$ light-years.

In a Universe with a flat geometry (i.e. $k = 0$), a feature of this size on the last-scattering surface would span an angle of about two degrees. A more precise calculation, based on the same effects, predicts that there should be strong temperature anisotropies on a scale of about one degree, rather than two, and thus accounts for the prominent peak in the angular power spectrum around $l = 220$. The other, lesser peaks in the angular power spectrum can be attributed to waves that have fully expanded once, recollapsed once, fully expanded twice and so on. The big peak in the angular power spectrum is often referred to as the Doppler peak, but it and the lesser peaks are all caused by essentially the same phenomenon and are sometimes collectively referred to as **acoustic peaks**.

At small angular scales (multipole numbers of 1000 or more), signs of the acoustic waves are expected to fade away. This is due to an effect called **Silk damping** and occurs where the acoustic waves would have such short wavelengths that they are effectively smeared out by the free movement of photons between encounters with charged particles. A similar effect connected with the finite sizes of molecules prevents sound waves of extremely short wavelength from travelling through air.

Now that the main influences on the angular power spectrum have been identified, we can ask how the angular power spectrum can be predicted on the basis of a given cosmological model (i.e. given values of H_0 , $\Omega_{\text{m},0}$, $\Omega_{\Lambda,0}$ and $\Omega_{\text{b},0}$ etc.) and some particular assumptions about the spectrum of density fluctuations (i.e. particular values of A and n_s). Making such a prediction is not easy. Fortunately for those who would like to do so, there are now publicly available computer programs such as ‘CMBFAST’, which can be used to calculate the shape of the angular power spectrum in a specified FRW model of the Universe. The ability to make such complicated predictions with relative ease is the key to using the observed CMB anisotropies to determine the main cosmological parameters, as is explained in the next section.

7.5.4 Determining cosmological parameters from the anisotropies in the CMB

Figure 7.26 shows a set of predictions for the shape of the CMB angular power spectrum based on fixed values for all the main cosmological parameters apart from $\Omega_{\text{b},0}$. The different curves in the figure show the effect of increasing the assumed current value of the baryon density parameter while keeping all the other parameters fixed. As you can see, the shape of the power spectrum changes significantly as $\Omega_{\text{b},0}$ is increased. Altering the other parameters would have different but analogous effects.

The general lesson to be drawn from Figure 7.26 is that the predicted shape of the angular power spectrum depends on the values of H_0 , $\Omega_{\text{m},0}$, $\Omega_{\Lambda,0}$, $\Omega_{\text{b},0}$, A , n_s etc. that are used in its calculation. This is an important point because it provides a method of deducing the values of all the cosmological parameters simultaneously. By comparing a large number of predicted power spectra with the available observational data regarding the power spectrum it is possible to identify a ‘best-fit’ spectrum, and hence

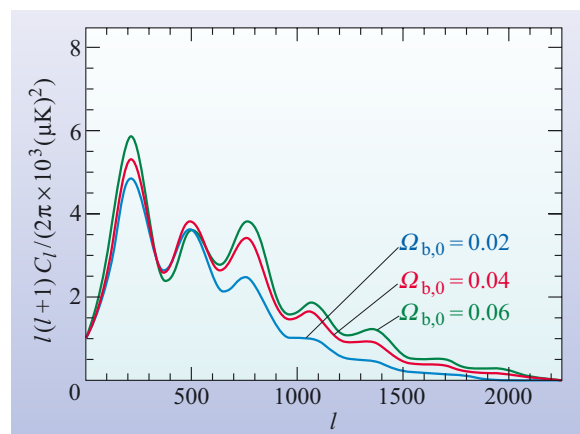


Figure 7.26 The angular power spectrum of CMB anisotropies, as predicted by big bang cosmology, for a range of values of the density parameter for baryonic matter, $\Omega_{\text{b},0}$. The differently coloured traces correspond to values of $\Omega_{\text{b},0}$ ranging from 0.02 to 0.06.

to come to some conclusions about the most probable values for the various cosmological parameters. More detailed analysis even allows uncertainties to be estimated.

This process of determining cosmological parameters from CMB anisotropy observations has already been carried out by a number of observational cosmology research groups. The latest values at the time of writing, those obtained by the WMAP team, are shown in Figure 7.27. Some of the parameters listed will be unfamiliar since they refer to aspects of cosmology that have not been discussed in this book. However from the list it is possible to draw the following values for the key cosmological parameters.

$$H_0 = (71 \pm 4) \text{ km s}^{-1} \text{ Mpc}^{-1}$$

$$\Omega_{\text{m},0} = 0.27 \pm 0.04$$

$$\Omega_{\Lambda,0} = 0.73 \pm 0.04$$

$$\Omega_{\text{b},0} = 0.044 \pm 0.004$$

Also included in the list are the values of some ‘secondary’ parameters that depend on the key parameters. Among them is a value for the age of the Universe which provides a valuable cross-check on the reasonableness of the ‘primary’ results. In the case of the WMAP findings, the favoured value for the age of the Universe is

$$t_0 = 13.7 \times 10^9 \text{ years}$$

Pleasingly, this does seem to be a realistic estimate for t_0 , since it makes the Universe older than the oldest stars we know of, the Population II stars in globular clusters. In view of this and other consistency checks the WMAP results have been well-received by cosmologists and are widely thought to represent the best numerical characterization of the Universe currently available.

7.5.5 Precision cosmology

The term **precision cosmology** was coined in 1996 by the American cosmologist Michael Turner. It expresses rather well the direction in which cosmology is thought to be heading. Gone are the days when cosmologists were free to speculate about the origin and evolution of the Universe with hardly any observational data to guide or fetter them. Cosmologists, like most other scientists, are now fenced in by large amounts of data, some of them very precise. Hubble’s original estimate of the constant that bears his name was wrong by a factor of about 10. The estimate based on WMAP’s anisotropy observations is thought to be within about 5% of the true value. Many of the other cosmological parameters are now thought to be known with a similar level of precision.

Studies of the CMB have certainly led the way into the era of precision cosmology and are expected to provide much of the data that will help to sustain it. A European space probe called *Planck*, planned for launch in 2007 or later, is expected to measure the CMB anisotropies on even finer scales than WMAP, and to determine the angular power spectrum almost as well as cosmic variance will permit. This will allow the cosmological parameters to be determined with even greater precision, perhaps with uncertainties of only 1% in some cases. However, the feeling among observational cosmologists that they are finally closing on the correct values of the cosmological parameters is not based on CMB measurements alone.

What has really boosted the confidence of observational cosmologists is finding that several independent lines of investigation are leading them towards the same

Description	Symbol	Value	+ uncertainty	– uncertainty
Total density	Ω_{tot}	1.02	0.02	0.02
Equation of state of quintessence	w	< -0.78	95% CL	–
Dark energy density	Ω_{Λ}	0.73	0.04	0.04
Baryon density	$\Omega_{\text{b}} h^2$	0.0224	0.0009	0.0009
Baryon density	Ω_{b}	0.044	0.004	0.004
Baryon density (cm^{-3})	n_{b}	2.5×10^{-7}	0.1×10^{-7}	0.1×10^{-7}
Matter density	$\Omega_{\text{m}} h^2$	0.135	0.008	0.009
Matter density	Ω_{m}	0.27	0.04	0.04
Light neutrino density	$\Omega_{\nu} h^2$	< 0.0076	95% CL	–
CMB temperature (K) ^a	T_{cmb}	2.725	0.002	0.002
CMB photon density (cm^{-3}) ^b	n_{γ}	410.4	0.9	0.9
Baryon-to-photon ratio	η	6.1×10^{-10}	0.3×10^{-10}	0.2×10^{-10}
Baryon-to-matter ratio	$\Omega_{\text{b}} \Omega_{\text{m}}^{-1}$	0.17	0.01	0.01
Fluctuation amplitude in $8h^{-1}$ Mpc spheres	σ_8	0.84	0.04	0.04
Low- z cluster abundance scaling	$\sigma_8 \Omega_{\text{m}}^{0.5}$	0.44	0.04	0.05
Power spectrum normalization (at $k_0 = 0.05 \text{ Mpc}^{-1}$) ^c	A	0.833	0.086	0.083
Scalar spectral index (at $k_0 = 0.05 \text{ Mpc}^{-1}$) ^c	n_s	0.93	0.03	0.03
Running index slope (at $k_0 = 0.05 \text{ Mpc}^{-1}$) ^c	$dn_s/d \ln k$	-0.031	0.016	0.018
Tensor-to-scalar ratio (at $k_0 = 0.002 \text{ Mpc}^{-1}$)	r	< 0.90	95% CL	–
Redshift of decoupling	z_{dec}	1089	1	1
Thickness of decoupling (FWHM)	Δz_{dec}	195	2	2
Hubble constant	h	0.71	0.04	0.03
Age of Universe (Gyr)	t_0	13.7	0.2	0.2
Age at decoupling (kyr)	t_{dec}	379	8	7
Age at reionization (Myr, 95% CL)	t_{r}	180	220	80
Decoupling time interval (kyr)	Δt_{dec}	118	3	2
Redshift of matter-energy equality	z_{eq}	3233	194	210
Reionization optical depth	τ	0.17	0.04	0.04
Redshift of reionization (95% CL)	z_{r}	20	10	9
Sound horizon at decoupling ($^{\circ}$)	θ_A	0.598	0.002	0.002
Angular diameter distance to decoupling (Gpc)	d_A	14.0	0.2	0.3
Acoustic scale ^d	ℓ_A	301	1	1
Sound horizon at decoupling (Mpc) ^d	r_s	147	2	2

^a from COBE (Mather *et al.* 1999)

^b derived from COBE (Mather *et al.* 1999)

^c $l_{\text{eff}} \approx 700$

^d $\ell_A \equiv \pi \theta_A^{-1}$ $\theta_A \equiv r_s d_A^{-1}$

Figure 7.27 The values of the various cosmological parameters obtained by comparing the WMAP observations with a range of cosmological models; CL indicates confidence limit. These values are taken from a paper by C. L. Bennett *et al.* and draw on the results obtained by many different observational cosmologists. (Note that some of these parameters have not been described in this book.)

conclusion. You saw evidence of this earlier, in the many different ways of trying to determine the key cosmological parameters. Some are inherently less accurate than others, but over the past few years they have been leading towards the same general conclusion. We live in a homogeneous and isotropic Universe with a flat ($k = 0$) spatial geometry and a scale factor $R(t)$ that is growing at an accelerating rate. The Universe is dominated by dark energy, but also contains a substantial amount of non-baryonic dark matter, a much smaller quantity of baryonic matter and an almost negligible amount of radiation.

Evidence concerning the composition of the Universe provides a good example of many different lines of investigation coming together. You saw in Section 7.4.1 that observations of Type Ia supernovae provide information about q_0 which is related to the combination $(\Omega_{m,0}/2) - \Omega_{\Lambda,0}$. The study of CMB anisotropies provides independent information and has tightly constrained $\Omega_{m,0} + \Omega_{\Lambda,0}$ since the time of BOOMERanG, and even more precise information has come from WMAP. Yet another line of argument involves the study of the large-scale distribution of galaxies that was briefly discussed in Chapter 4. The three-dimensional distribution of galaxies can also be described by a power spectrum, although it is very different from the angular power spectrum of the CMB. Nonetheless, the galaxy power spectrum contains features that limit the plausible range of values for $\Omega_{m,0}$. Although it is the study of CMB anisotropies that provides the most precise values for these density parameters, it is the agreement and combined effect of all the results that persuade many cosmologists that they are on the right track.

Cosmology is a speculative field of enquiry, but the constraints and tests provided by observational cosmology ensure that the core of that enquiry will be disciplined and effective. The new era of precision cosmology will be challenging, but it is safe to predict that it will also be exciting.

7.6 Summary of Chapter 7

Measuring the Hubble constant, H_0

- H_0 measures the current rate of expansion of the Universe.
- H_0 is traditionally determined by means of the Hubble diagram, a plot of redshift against distance for distant galaxies. When making such a plot the redshift and the distance must be determined independently. The Hubble constant is obtained from the gradient of the plotted line.
- The HST Key Project team used Cepheid variables to calibrate five other methods of distance measurement. Using these together with independent measurements of galaxy redshifts they concluded that $H_0 = 72 \pm 8 \text{ km s}^{-1} \text{ Mpc}^{-1}$.
- Other methods of determining H_0 include those based on gravitational lensing (via time delays between fluctuations in different images of the same lensed galaxy) and those based on observations of anisotropies in the cosmic microwave background radiation (CMB).

Measuring the current value of the deceleration parameter, q_0

- q_0 measures the current rate of change of cosmic expansion.
- q_0 may be determined from the Hubble diagram by observing the curvature of the plotted line at $z > 0.2$, but early attempts to do this were highly inconsistent.

- Results obtained using Type Ia supernovae as distance indicators suggest that q_0 is negative, implying that the expansion of the Universe is speeding up. It has become traditional to express the value of q_0 in terms of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$, using the relation

$$q_0 = \frac{\Omega_{m,0}}{2} - \Omega_{\Lambda,0}$$

Measuring the current values of the density constants, $\Omega_{\Lambda,0}$, $\Omega_{m,0}$ and $\Omega_{b,0}$

- The current values of the density parameters $\Omega_{\Lambda,0}$, $\Omega_{m,0}$ and $\Omega_{b,0}$ measure the densities associated with the cosmological constant (dark energy), matter of all kinds and baryonic matter, relative to the critical density $\rho_{\text{crit}} = 3H_0^2/8\pi G$.
- Results based on observations of anisotropies in the CMB strongly favour $\Omega_{\Lambda,0} + \Omega_{m,0} = 1$.
- When these are combined with the results of observations of Type Ia supernovae, favoured values of $\Omega_{\Lambda,0}$ and $\Omega_{m,0}$ are typically $\Omega_{m,0} = 0.30 \pm 0.10$ and $\Omega_{\Lambda,0} = 0.70 \pm 0.10$.
- The current value of the density constant for baryonic matter can be determined in a number of ways. This quantity is constrained by primordial nucleosynthesis calculations which, if they are to agree with observations, require $0.02 \leq \Omega_{b,0} \leq 0.05$.
- Direct assessments of $\Omega_{b,0}$, based on baryon inventories, are beset by many uncertainties, but generally favour lower values: $0.007 \leq \Omega_{b,0} \leq 0.041$.

Anisotropies in the CMB and precision cosmology

- Although highly isotropic, the CMB exhibits anisotropies in intensity at the level of a few parts in 100 000 over a range of angular scales. These can be mapped, and are usually shown as variations in the temperature of the CMB.
- The angular power spectrum of an anisotropy map shows the level of variation that is present on any specified angular scale (or, equivalently, the angular power at multipole number l).
- Values of cosmological parameters may be extracted from anisotropy measurements by comparing the observed angular power spectrum with that predicted by big bang cosmology. Recent results from the WMAP space probe indicate $\Omega_{\Lambda,0} = 0.73 \pm 0.04$, $\Omega_{m,0} = 0.27 \pm 0.04$, and $\Omega_{b,0} = 0.044 \pm 0.004$, implying a Universe dominated by dark energy, and in which most of the matter is non-baryonic dark matter.
- WMAP measurements also indicate that $H_0 = (71 \pm 4) \text{ km s}^{-1} \text{ Mpc}^{-1}$, and that the age of the Universe is $t_0 = 13.7 \times 10^9$ years.
- The results may indicate that we are now entering an era of precision cosmology in which cosmological speculations will be tightly constrained by measurements, and quantities that were previously very uncertain will become accurately known.

Questions

QUESTION 7.9

Using a variety of measurements (not just CMB), outline the observational basis of the claim that

- (a) the Universe is dominated by dark energy
- (b) most of the matter is dark
- (c) most of the dark matter is non-baryonic.

QUESTION 7.10

On the basis of the very incomplete account given in this chapter, outline the role that space technology has played in observational cosmology and the role that it is expected to play in the future development of the subject.

QUESTION 7.11

Quote some examples to show the importance of terrestrial (as opposed to space-based) observations in cosmology.
